Explaining the MLB Home Run Record of 2019 with Quality of Pitch (QOP™)[1]

by: Jason Wilson[2], Joseph Lane[2], and Wayne Greiner[1]

Summary: After establishing correlation between quality of pitch and home runs, we determined that quality of pitch accounts for a 26% to 40% of the amount of the variation in the proportion of home runs in MLB. Two of the six pitch components under study changed from historic levels in 2019: horizontal break and location. On average, pitchers are moving their pitches from the middle of the strike zone (easier to hit home runs) to low and close to the batter (outside the strike zone, where it is harder to hit home runs). Nevertheless, pitches in 12 out of 13 locations (43 out of 52 when handedness splits are considered) experienced an increase in the proportion of home runs from 2018 to 2019. These changes are consistent with pitchers reacting to a perceived threat of increased home runs whether due to ball changes, batter approach, or otherwise. They are also consistent with pitches flying unintentionally straighter due to balls with less drag. The statistics do not indicate whether the explainable proportion of increased home runs is caused by poorer pitching or is causing poorer pitching. Regardless, the quality of pitch in 2019 is projected to finish at a record low, while home runs, at a record high. If pitchers are attempting to move their pitches to safer locations, this strategy generally appears to be backfiring. While pitch quality is one of the factors in the home run surge of 2019, it offers only a partial explanation, according to the correlations.

Note: The data used for this paper were through June 18, 2019, which were 443,127 pitches, around 60% of the season's pitches. Although the specific numbers at the end of the season will vary from those shown in this report, it is expected that the trends, and therefore the conclusions, will remain the same. Should a surprise occur and overturn any conclusion of this report, we will update it. Otherwise it will remain as it is, due to the amount of effort involved in revising all of the analyses shown.

---

[2] Biola University, 13800 La Mirada, CA 90639; jason.wilson@biola.edu; joseph.a.lane@biola.edu

Table of Contents

1. Correlation between QOP and HR 3
 1.1 QOPA and home runs by year 3
 1.2 Plot of HR proportion as a function of QOPA 4
 1.3 Scatterplots by pitch type 5
 1.4 Generalized linear logistic regression models, within year 9
 1.5 Generalized linear logistic regression model, including year 11
2. How much of a factor is quality of pitch in HR%? 12
3. What Changes in Pitching are Related to Home Run% in 2019? 13
 3.1 Changes in Pitch Components 14
 3.2 The Strike Zone Model 17
 3.3 How Location Changed 17
4. Conclusions 25
Appendix A: Coefficients and p-values from General Linear Models by year 26
Appendix B: Generalized Linear Models, by pitch type, with year as a factor 29
 Appendix B.1. All pitches 29
 Appendix B.2. Only pitches swung at 35
Appendix C: Generalized Linear Models, by pitch type, without year as factor 40
Appendix D: Generalized Linear Models, by pitch type, with year as numeric 44
Appendix E: Control Charts showing changes in components for pitches swung on 49
Appendix F: Comparison of our zone numbers with PITCHf/x and Statcast 50
Appendix G: Pitcher-Batter Handedness Splits 52
Appendix H: Control Charts for Pitcher-Batter Handedness Splits 58

The purpose of this technical report is to document the research we undertook in July 2019 to understand the relationship between pitching and the home run surge in Major League Baseball in 2019. The discussion of whether it is the ball, or the batters, or even other factors, is widely discussed in the media. We covered the major sources current as of 2018 in our previous work[3]. In this report we focus exclusively on the data. It differs from our previous work in two primary respects. First, the impetus is the even greater surge of 2019[4]. In this study, the data is current as of July 18, 2019, with over 400,000 pitches. Second, our earlier paper focused primarily on 2017, while here we do more analysis on 2008 to 2019, in order to gain a greater historical perspective.

The paper is organized as follows. Section (1) shows the correlation between quality of pitch and home run proportion from several angles. Section (2) attempts to quantify the amount of correlation between quality of pitch and home run proportion. Having established that a relationship exists, Section (3) addresses the main questions – Did the pitching change in 2019? If so, how? Conclusions are drawn in Section (4). Because this is a lengthy technical report, and some of the regression models are tedious, the reader is advised to use the Table of Contents and read selectively, focusing on the portions of greatest interest.

1. Correlation between QOP and HR

There is a relationship between QOP and home runs (HR). We establish this with four lines of evidence:

(i) scatterplots of QOPA by HR,
(ii) a graph of the functional relationship between QOPA and HR,
(iii) within-year cross-validated general linear models which successfully predict HR, and
(iv) an across year explanatory general linear model.

We take each of these four in turn.

Since QOP depends on six different pitch components, different pitch types have different QOP averages (QOPA). Therefore, throughout this paper we will often look at results by pitch type[5]. In particular, we will focus on the six most common pitch types, change up (CH), curveball (CU), four seam fastball (FF), two seam fastball (FT), sinker (SI), and slider (SL). Throughout this paper, the six components of QOP have been abbreviated as: rise, breakpt (breaking point), tot.brk (total break = vertical break), h.brk2 (horizontal break), loc (location, on our unitless scale), and start.speed (MPH at 50’, or about 5’ from release point). See earlier papers for an explanation of the QOP model and the components.

1.1 QOPA and home runs by year

The relationship between the number of home runs vs. quality of pitch average (QOPA) can be seen in Table 1. Glancing at the numbers shows that QOPA has been in a pretty narrow channel from 2008 to 2018, 4.46 to 4.59. , home runs tend to increase with increased QOPA, rather than decrease, as expected. As will be shown below, this is due to the blending together of different variables that, when identified and separated, reveal the correlation.

---

[3] Wilson, Jason; Jordan Wong, Jeremiah Chuang, Wayne Greiner. Explaining the MLB Home Run Record of 2017 with QOP. Technical Report. 2018.

[4] See, for example, https://theathletic.com/1044790/2019/06/25/yes-the-baseball-is-different-again-an-astrophysicist-examines-this-years-baseballs-and-breaks-down-the-changes/.

[5] We explored scaling all pitch types onto the same scale. Surprisingly, it had very little effect on the rankings of pitchers by QOPA. The downside of such scaling is that calculated QOPA’s were not only a linear combination of the six pitch components (rise, total break, vertical break, horizontal break, location, and speed), but an additional scale factor had to be applied, which would need to be calculated per season. In practice, allowing them to simply “fall where they land” has provided insight into the relationship between the quality of different pitch types. Therefore, at this time we have considered the added complexities of interpretation, and distance from the original components, to not be worth the small gain of having QOPA’s with the same scale.

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019?
HR	4878	5042	4613	4552	4934	4661	4186	4909	5610	6105	5585	**6500?**
% inc	0.98	1.03	0.92	0.99	1.08	0.95	0.90	1.17	1.14	1.09	0.91	**1.16?**
QOPA	4.47	4.51	4.46	4.47	4.57	4.57	4.57	4.58	4.59	4.49	4.48	**4.40?**

Table 1. Home runs vs. QOPA. The 2019 figures are a conservative projection for the home runs; the QOPA is current through July 18, 2019

1.2 Plot of HR proportion as a function of QOPA

In order to identify the first variable to separate and reveal the correlation between home runs and QOP, consider the proportion of home runs as a function of QOPA. After grouping the QOPVs into bins of -0.5 to 0.5, 0.5 to 1.5, 1.5 to 2.5, …, 7.5 to 8.5, and 8.5+, we computed the proportion of home runs in each bin and plotted it against the center of the bin. E.g. for the bin 2.5 to 3.5, we plotted QOPA=3 vs. HR=0.010 for all pitches (red line in Figure 1).

There is a clear functional relationship between QOPA and home runs. In particular, for QOPV between 0 and around 3 (the poorest quality pitches), the home runs actually increase as QOPA increases. This is because pitches on the low end of the 0-3 range tend to be the chaser, outside, ball in dirt, etc. pitches, but become more hittable as their QOP rises. Then, for the main pitches of interest, QOPV around 3 and up, there is a clear decrease in HR proportion as QOPV increases. This holds for all pitches (red curve) as well as pitches swung at (blue curve). The difference is that there is necessarily a higher proportion of home runs on pitches swung at.

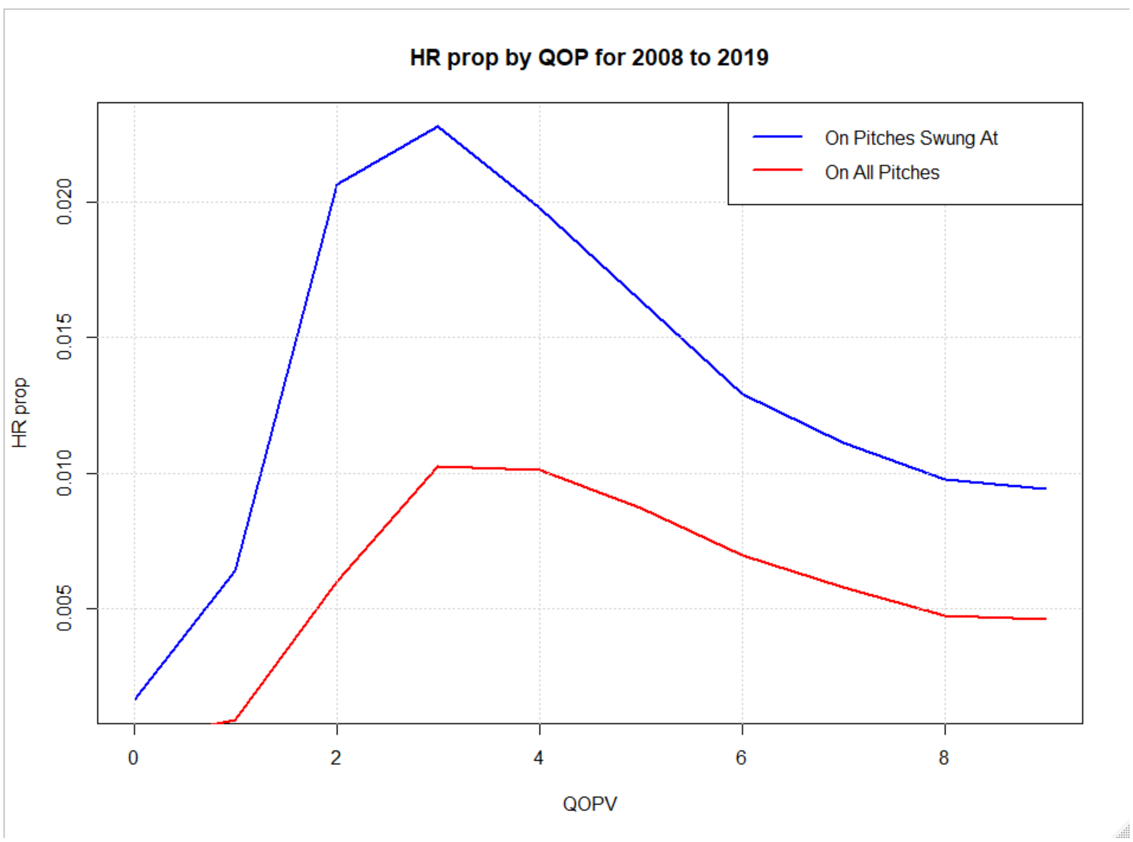


Figure 1. Home run proportion as a function of Quality of Pitch Values (QOPV)

Now, plotting the home run proportion by QOPA in two groups, above and below 3, reveals the correlation (Figure 2). As can be seen, there is a positive correlation between QOPV and home runs for low quality pitches (QOPV<=3) and a negative correlation for mid to high quality pitches (QOPV>3). Thus, there are two different categories of QOP values that relate to home runs. This observation will be important later.

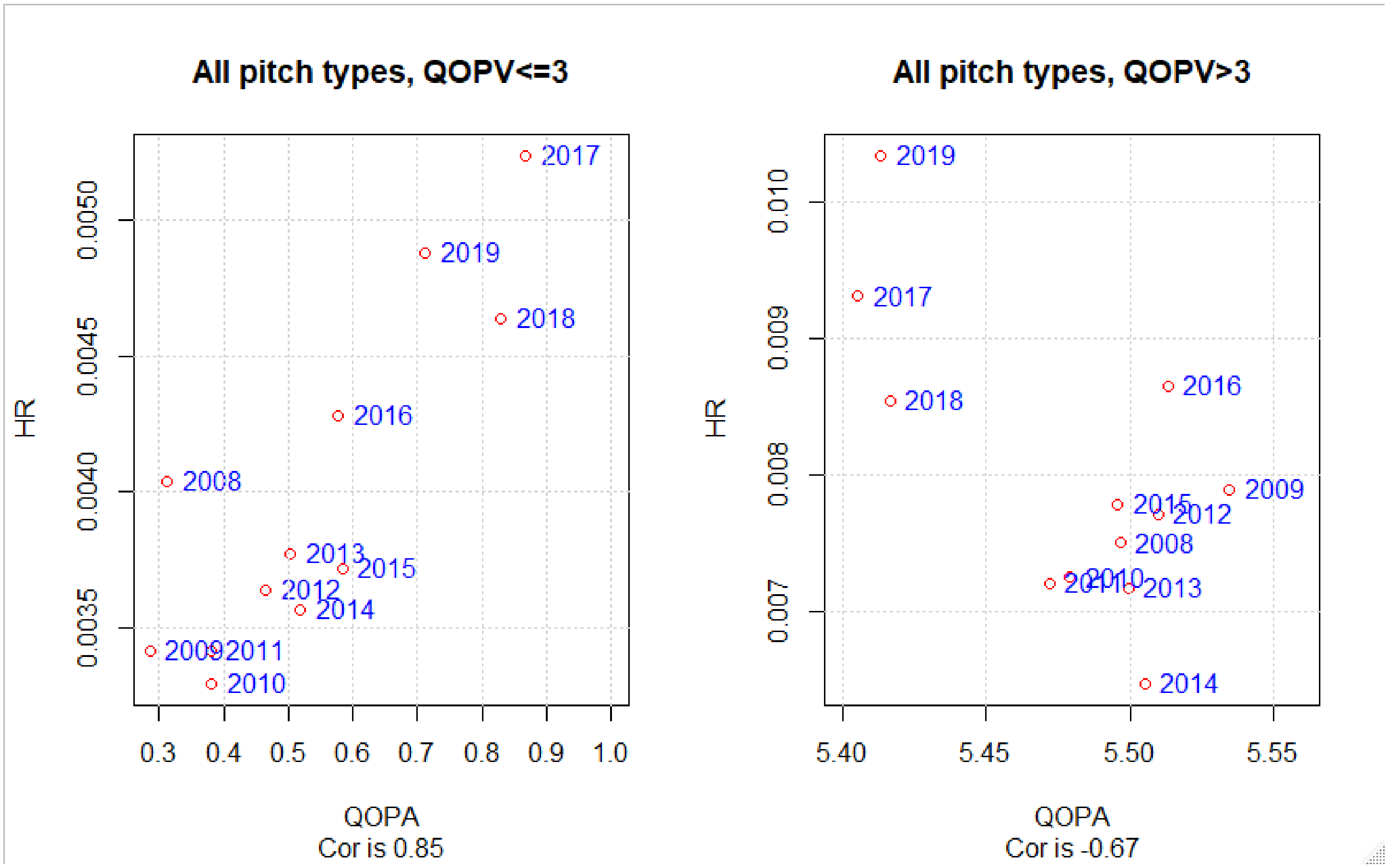


Figure 2. Plots of home runs (HR) by QOPA, split according to pitches with QOPV <=3 versus QOPV > 3

1.3 Scatterplots by pitch type

In this subsection we examine scatterplots of QOPA vs. HR proportion by pitch type. We look at the numbers in three different ways: all pitches (Figure 3), split by QOPV<=3 and QOPV>3 (Figures 4 & 5), and only pitches swung at (Figure 6). Since we are considering the best intended contact hits (home runs), we narrow the set of pitches to those swung at, or where the batter intended contact. The overall pattern for the three cases is similar (excepting QOPV<=3).

We predicted negative correlation for all cases – as QOPA increases, the proportion of homeruns decreases. For all three cases, the same basic observation holds: there is strong negative correlation for curveballs (CU), moderate negative correlation for four seam fastballs (FF) and sliders (SL), and weak negative correlation for two seam fastballs (FT) and sinkers (SI). There is positive or no correlation for change-ups (CH). These results are perhaps not surprising since the original QOP conception was for curveballs, and the one-size-fits-all QOP model rates change-ups higher QOP for higher speed, which is contrary to the purpose of change-ups[6].

[6] This is perhaps the weakest point of our model. We have considered changing the model to optimize it for each pitch type, particularly change-ups. This would undoubtedly result in increased precision, including for home run prediction. The downside is that such a move would complicate the model beyond our preferred level. In its present form, the model is simple and completely transparent, up to the lack of public disclosure of the model coefficients. All pitches are held up to the identical measuring stick of QOPV, with its six components, and may be directly compared with one another. For this reason, we continue to focus our

Omitting the change-up, then the mean correlation of the five pitch types for all pitches is -0.51 and -0.52 for pitches with QOPV >=3. These correlations give a coefficient of determination around 26%.

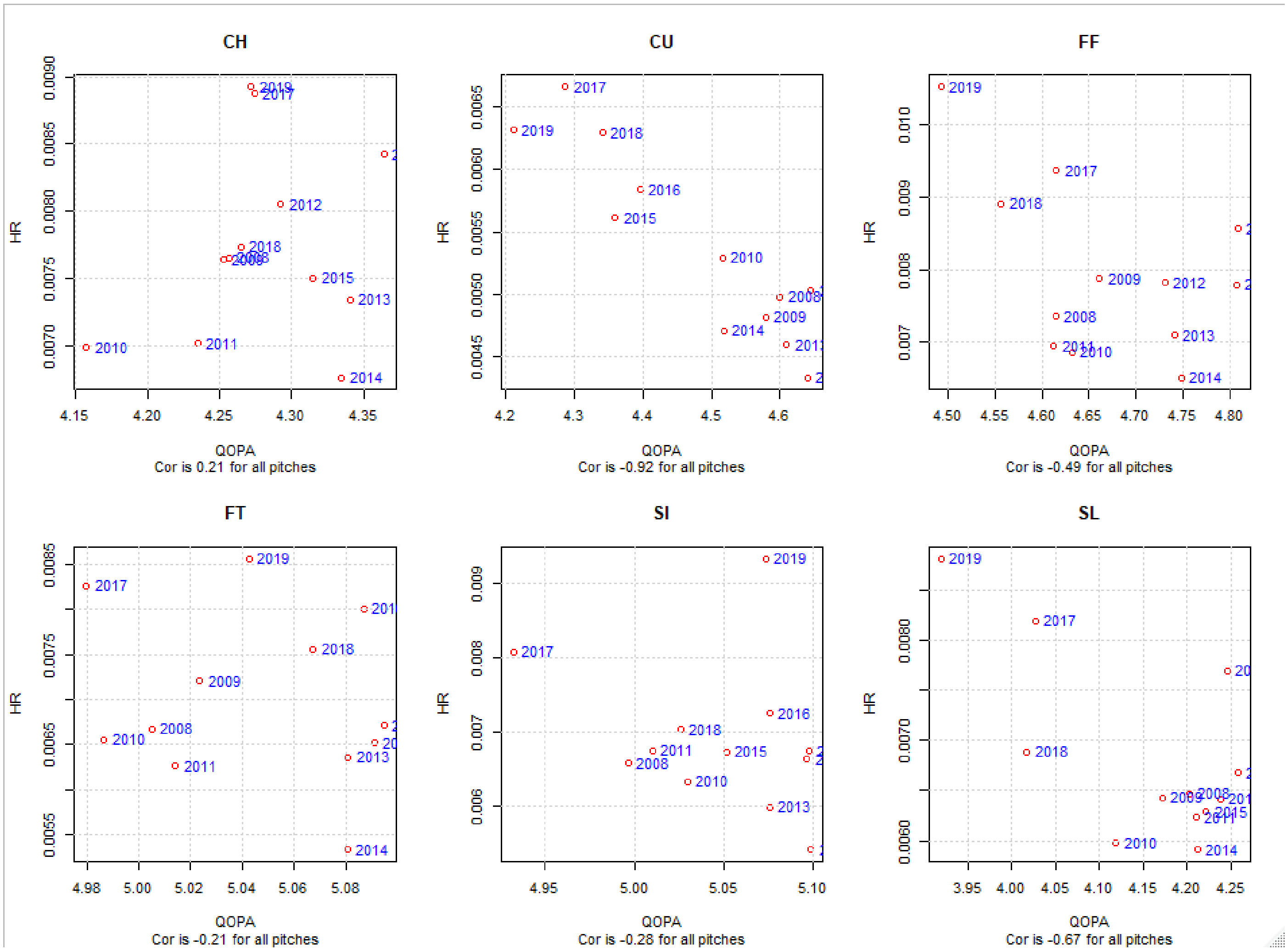


Figure 3. Scatterplot of QOPA vs. HR proportion for all pitches, by pitch type.

development efforts in the direction of simplicity of interpretation at the expense of added precision and power. This reflects our position as public analysts, as opposed to MLB analysts for a particular team.

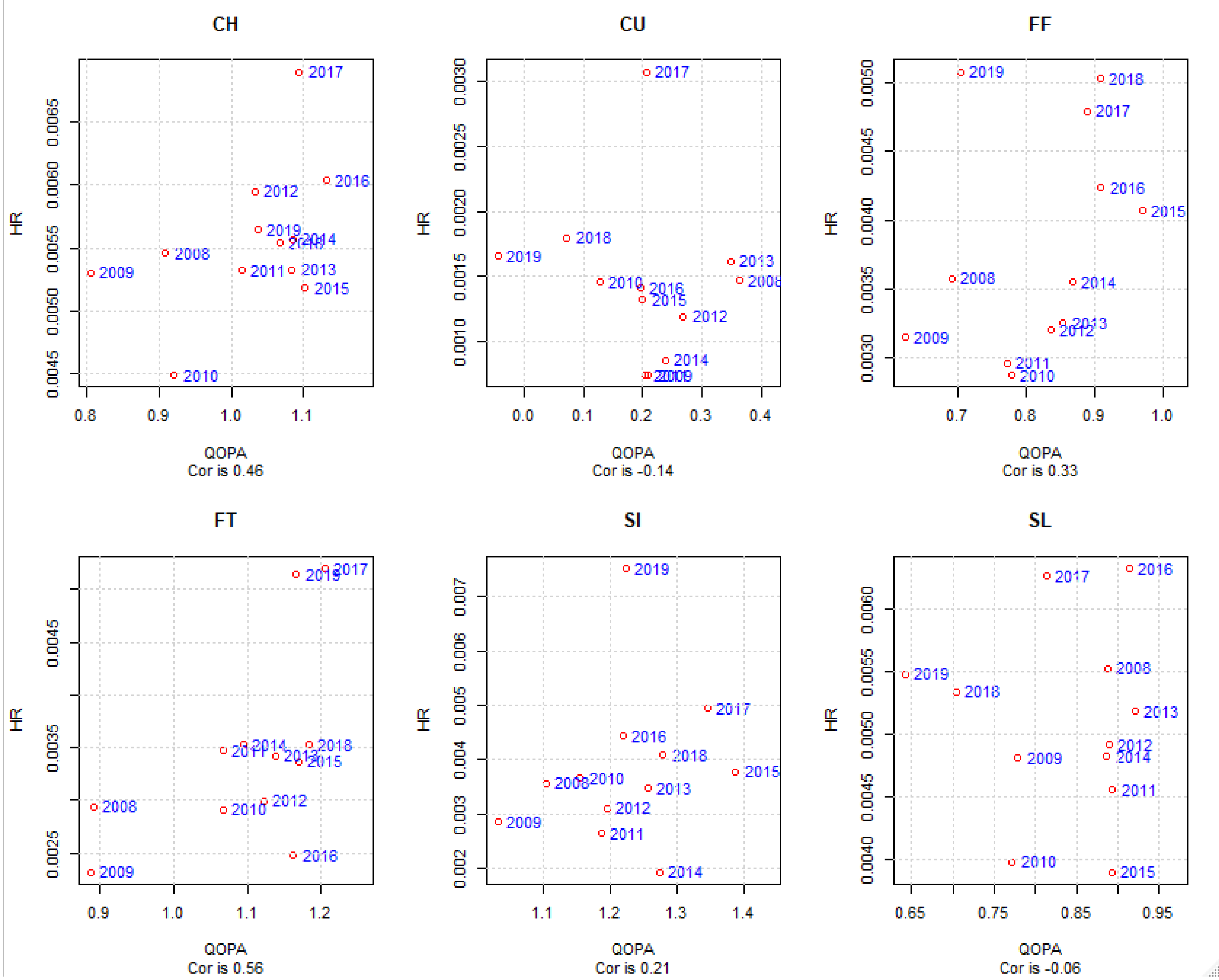


Figure 4. Scatterplot of QOPA vs. HR proportion for only pitches with QOPV<3, by pitch type.

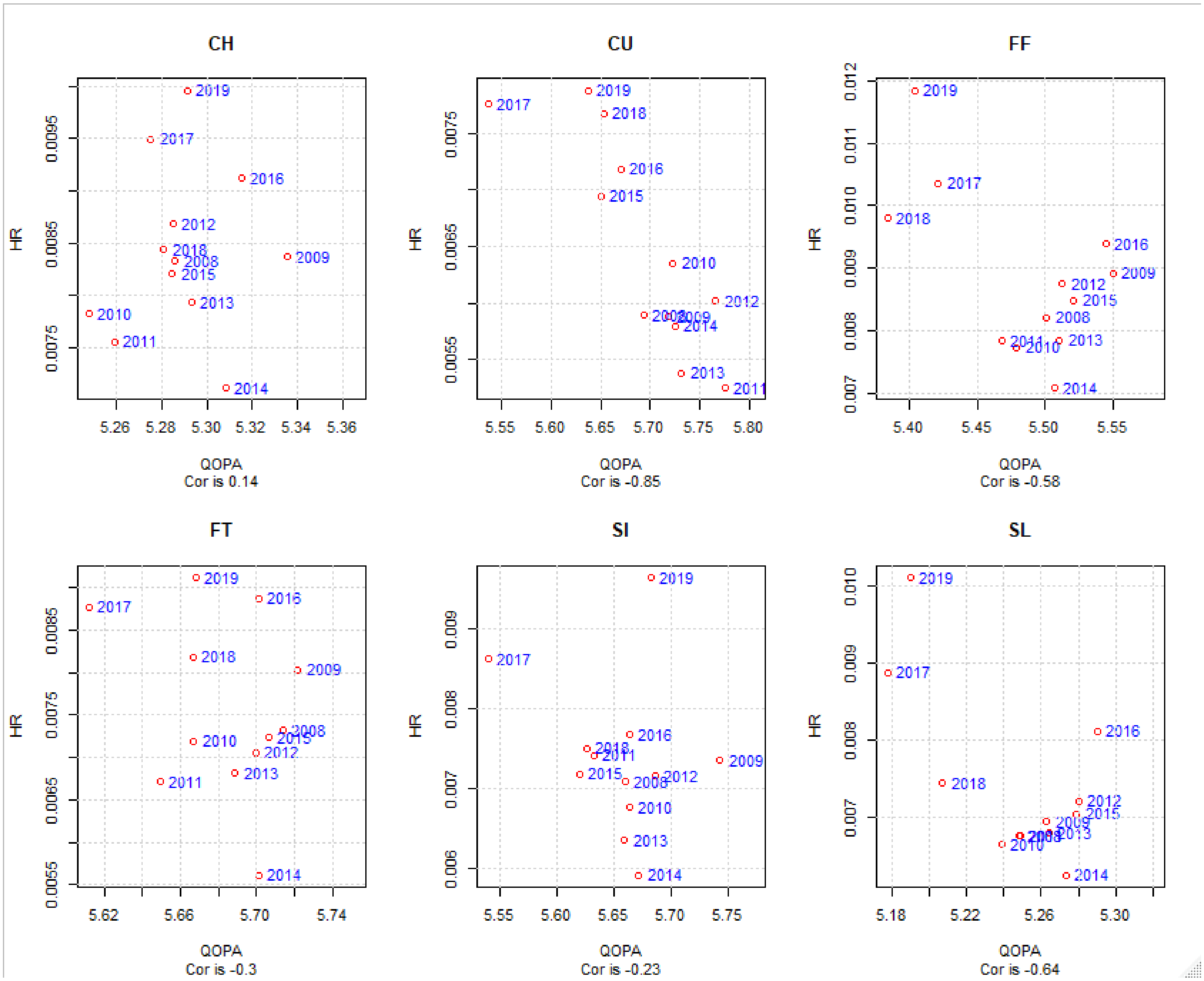


Figure 5. Scatterplot of QOPA vs. HR proportion for only pitches with QOPV>=3, by pitch type.

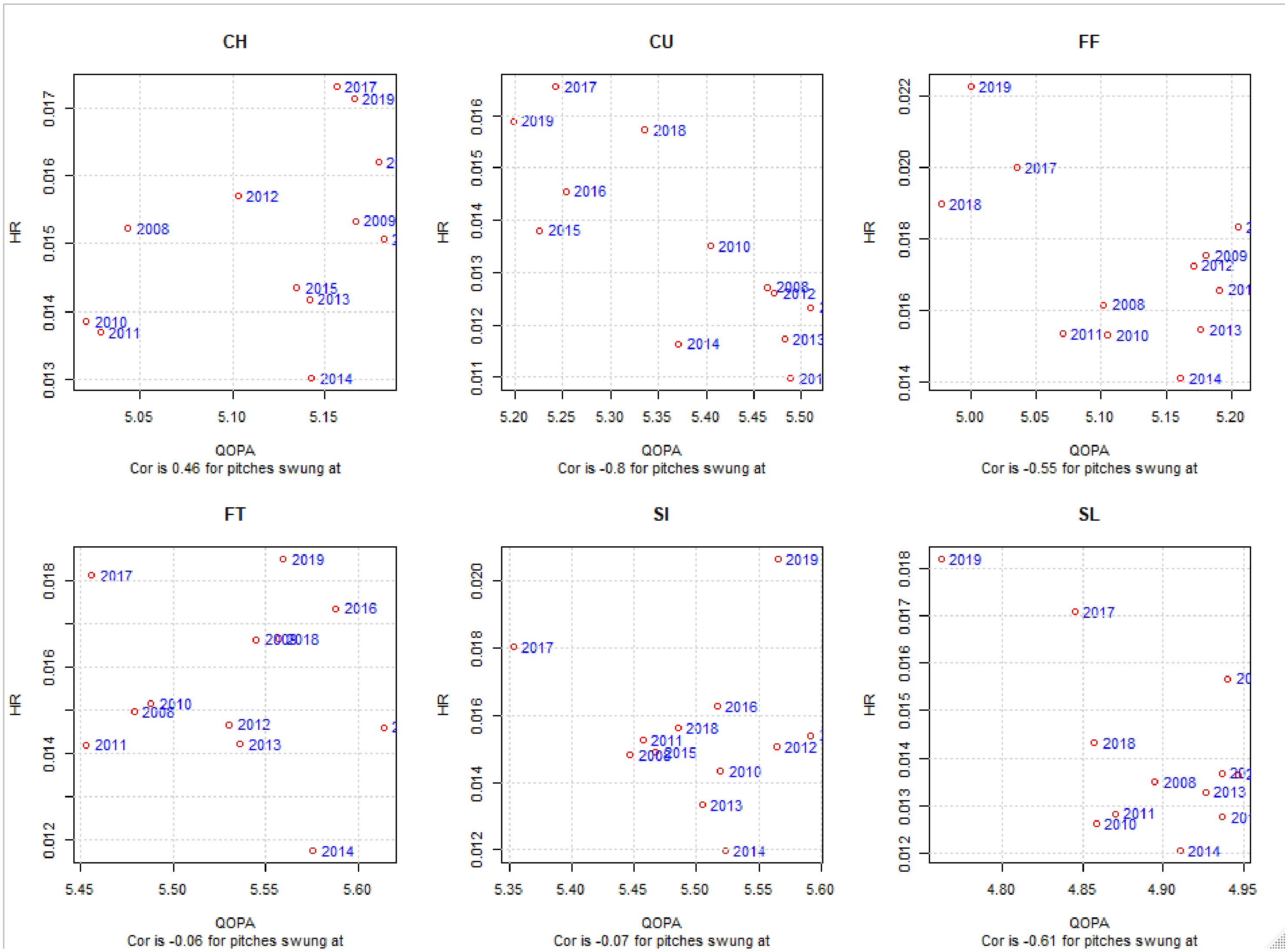


Figure 6. Scatterplot of QOPA vs. HR proportion for only pitches swung at, by pitch type.

1.4 Generalized linear logistic regression models, within year

In our 2018 paper we did a detailed study of the 2017 home run surge which included exploring generalized logistic regression models. One of the conclusions was that batter height (representing change in batters) and handedness match-up were significant enough to need inclusion in order for the models to be predictive. Therefore, our first models considered only the six QOP components, and generated one model per pitch type. Following the reasoning of our previous work, we used the following model, where HR% denotes the proportion of home runs:

$$HR\% = intercept + rise + breakpt + tot.brk + h.brk2 + loc + start.speed$$

Algorithm 1 was used to generate the results.

1. Remove pitches with QOPV < 0 2. For each pitch type, randomly divide the pitches of one season into 1/2 test and 1/2 validation datasets 3. Construct a logistic regression model from the test data and apply the model to the validation data [for each pitch type] 4. Sum the explained HRs across pitch types 5. Construct a confidence interval for actual number of HRs 6. Set flag=1 if prediction is in CI, flag=0 if outside 7. Repeat steps 1-6 1000 times 8. Validation % = sum of flag divided by 1000
Algorithm 1. Logistic regression model, run 1000 times with 50% of 2017 data as test sample to construct model and 50% as validation sample.

The reason that more than one model is constructed is that if all of the data is used to build the model, it can fit the data perfectly. Therefore, we used one standard approach, which is to use half of the data to build a model which is used to explain the results in the other half of the data[7]. We built models from 1000 different random test and validation sets. The reader may view a set of sample models for 2017 in our previous paper. For this research we built the models for all years, totaling of 12,000 models (Table 2)[8]. For example, for 2017, the final result was that 83.7% of our test set models accurately explained their validation set model. By "accurately explained", we mean that the predicted number of home runs fell within a 95% confidence interval of the validation home runs[9]. For comparison, we show the results for 90%, 95%, and 99% confidence intervals (Table 2).

Confidence Level	**90%**	**95%**	**99%**
2019	0.714	0.796	0.882
2018	0.749	0.828	0.934
2017	0.768	0.837	0.933
2016	0.746	0.824	0.922
2015	0.759	0.828	0.936
2014	0.759	0.835	0.925
2013	0.764	0.836	0.938
2012	0.750	0.817	0.915
2011	0.760	0.827	0.933
2010	0.742	0.824	0.922
2009	0.737	0.825	0.938
2008	0.744	0.823	0.933

Table 2. Validation rates for home run models.

[7] This is an explanatory model, in that it explains the results of the data from 2008 to 2019, and it is statistically valid. Another approach is to build a model using all of the data from one season and use that model to predict the home runs of the subsequent season. We tried that, but the only two successful predictions, i.e. predicted number of home runs within the 95% confidence interval of the subsequent season's home runs, was 2010-2011 and 2016-2017. Thus, while the full model for 2016 does successfully predict 2017 home runs, we did not consider it to be validated statistically because the same technique only worked for 2 out of 9 season pairs. More detailed interpretation of these models can be found in Wilson, et al. 2017.

[8] It turns out that two of the years actually had only 902 models, instead of 1,000, due to an index set to i=2, for debugging purposes. The missing 2*98=196 models will not appreciably change the results of Table 2.

[9] The confidence interval was generated using R's prop.test() function, and multiplying it by the number of pitches in the validation set. The prop.test() function uses a score test for its confidence interval, which is close to the common Wald or Agresti-Coull confidence interval for proportions.

The meaning of Table 2 is that the pitch components are sufficient to explain the record number of home runs within a particular year. This does not rule out other factors, because as with any model there is error (the projected number of home runs is not exact). However, it does provide very strong evidence that the pitch components are related to home runs.

Not only are pitch components related to home runs, the actual model performance is strikingly similar across 2008 to 2018, and then it drops a bit in 2019. The difference is interesting, but the models themselves do not offer an explanation for the change in behavior because they are created within each year. This leads to our next analysis.

1.5 Generalized linear logistic regression model, including year

While it is nice to have a model which explains the data for a particular year, we next combined the data for all years, giving the following model (HR% denotes proportion of home runs):

HR% = intercept + rise + breakpt + tot.brk + h.brk2 + loc + start.speed + Year

This is identical to the previous model, except *Year* is added as a categorical factor. This makes a different mean effect for each year, without requiring a linear trend across years[10]. There are six models, one for each pitch type. To highlight the quality of the model, see Table 3, the output for the sinker:

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.610317   0.439576   3.663 0.000249 ***
year2009     0.076492   0.055584   1.376 0.168775
year2010    -0.015668   0.056024  -0.280 0.779740
year2011     0.035694   0.057353   0.622 0.533715
year2012     0.041283   0.059827   0.690 0.490173
year2013    -0.058467   0.065452  -0.893 0.371705
year2014    -0.177045   0.066696  -2.655 0.007943 **
year2015     0.024271   0.064411   0.377 0.706305
year2016     0.127134   0.060443   2.103 0.035433 *
year2017     0.250132   0.063625   3.931 8.45e-05 ***
year2018     0.089964   0.063113   1.425 0.154032
year2019     0.411301   0.069343   5.931 3.00e-09 ***
rise         0.481266   0.304681   1.580 0.114204
breakpt     -0.083128   0.009342  -8.898  < 2e-16 ***
tot.brk     -0.289201   0.017796 -16.251  < 2e-16 ***
h.brk2      -0.418483   0.048601  -8.611  < 2e-16 ***
loc         -0.099948   0.008914 -11.213  < 2e-16 ***
MPH         -0.054897   0.004724 -11.620  < 2e-16 ***
---
    Null deviance: 68139  on 842019  degrees of freedom
Residual deviance: 67413  on 842002  degrees of freedom
AIC: 67449

[1] "Predicted HRs:  5683 Actual HRs: 5685"
```

Table 3. Generalized Linear Regression Model Output: Sinker with Year as categorical factor

All of the pitch components are highly statistically significant, with the exception of *rise*, which is not surprising for this pitch type. The sign of all pitch coefficients are as expected, except for *loc*, which will be discussed next. Regarding years, we see a negative coefficient in 2014 and significant positive coefficients in 2016, 2017, and 2019. Indeed, 2014

[10] Initially, we added *Year* as a numeric variable (Appendix D), however, we realized that modeled *Year* as a continuous variable, which turned out to have a positive slope. This indicated that HR% increased over the years, which is certainly true due to 2016, 2017, and 2019, not what we wanted to look at for *Year*.

was the record low number of home runs in the data while 2016, 2017, and 2019 are record highs. The model predicts 5683 home runs off of sinkers, whereas there were 5685. These confirm the model is behaving as expected.

Regarding the negative sign for the location coefficient, *loc* represents the distance away from the corners of the strike zone. Therefore, we would expect a positive coefficient, which would mean that the larger the location number, the further from the corners of the strike zone, and therefore the more hittable. The same negative sign occurss for the within-year models as well (Appendix A). This is the only surprise for the model coefficients in the models, being consistent throughout. The reason is that there are 8,363,508 pitches in the dataset[11], but with only 3,858,843 pitches swung at. This is 46% of the pitches. Of the 54% which were not swung at, many were due to poor location and therefore declined by the batter, meaning that larger *loc* scores for *all pitches* predict no swing and therefore no home run. In confirmation of this, we re-ran the model on only the 3,858,843 pitches swung at and in every case the *loc* coefficient was positive (Appendix C). Besides this, the same trends occur between either set of models. Similar results would occur if we ran the models on pitches with QOPV>3, the previously discovered split point.

In summary, we have provided five different lines of evidence to establish a correlation between quality of pitch and home runs. In particular, we showed that simple scatterplots of annual QOPAs by HR% reveals correlation with all major pitch types except the change-up, which is due to its technique. When HR% is viewed as a function of QOPA, a sort of quadratic shape emerges. This gives rise to a split point where pitches with QOPV >=3 exhibit the expected negative correlation. Digging deeper, both within and between year cross-validated generalized linear regression models successfully predict HR% and have highly significant pitch components.

Having established that pitch quality is related to home runs, one may ask –How much of a factor is quality of pitch in HR%? What changes in pitching are related to changes in HR% in 2019?

2. How much of a factor is quality of pitch in HR%?

Having established a relationship between quality of pitch and home runs, the next question is how much of a factor is it? Other plausible factors include:

(i) ball
(ii) batter approach
(iii) other

Some of the ideas and literature behind the first two were discussed in Wilson et. al. (2018). In this section, we attempt to estimate the proportion of influence of pitch quality.

The conventional approach for determining how much of the result a model explains is R^2. From the scatterplots of QOPA vs. home run % from section (1), we obtained a rough coefficient of determination, or R^2, of 26%. Do we get a match from the regression models? For logistic regression models, only a pseudo-R^2 is available[12], which we have used[13]. The pseudo-R^2 numbers for the models including the six pitch components + *Year* and can be viewed in Appendix B, while the models without *Year* are in Appendix C. The summarized results are below in Table 4.

---

[11] As of this writing, which has pitches through July 18, 2019. The dataset will increase by approximately 300,000 pitches by the end of 2019.

[12] See https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/ for a listing of the statistics and https://statisticalhorizons.com/r2logistic for a good discussion.

[13] Different attempts have been made. For these models, there is a wide variation in the results. As a principled method, we took the mean of the eight different methods given by the *PseudoR2* function in R's *DescTools* library. There are actually nine given, but two are McFadden and McFadden-Adjusted, so we will use only McFadden-Adjusted. It turns out that McFadden gives one of the highest results, so this substantially lowers the overall estimate. The pseudo-R^2 used are: McFadden, Cox & Snell, Nagelkerke, Aldrich & Nelson, Veal & Zimmermann, Effron, McKelvey & Zavoina, and Tjur. The reason for taking the mean of the results is that

Pitch Type	Mean pseudo-R^2 with year	Mean pseudo-R^2 without year	Min pseudo-R^2	Max pseudo-R^2
CH	0.453	0.453	0.002	0.930
CU	0.486	0.486	0.002	0.962
FF	0.269	0.268	0.001	0.653
FT	0.424	0.423	0.001	0.918
SI	0.441	0.441	0.001	0.937
SL	0.407	0.407	0.002	0.888

Table 4. Pseudo-R^2's for Generalized Linear Logistic Regression Models. Model output for "Mean pseudo- R^2 with year" are in Appendix B and for "Mean pseudo-R^2 without year" are in Appendix C.

All of the mean pseudo-R^2's are in the 40% range, except for the four seam fastball (FF) at 26.9%. These are "in the ballpark" of the 26% obtained from the scatterplots. Note, however, that the scatterplots only used the summarized QOPA by year, whereas the generalized linear logistic regression models used the six QOP components separately, and all pitches. It should be expected that the R^2 would be higher.

This is higher than the proportion of variation estimated in our previous study[14]. The reason is that only a single season was considered, whereas here we consider twelve seasons of data[15]. As seen in Table 3, *Year* does not account for much of the variation[16], which is consistent with many of the *Year* coefficient p-values either non-significant, or substantially less significant than the pitch components (Appendix B). **Therefore, the six pitch components do explain effectively all of the variation explained by the model, which we estimate around 40%. Thus, it is plausible to say that the six pitch components account for around 40% of the variation in home run proportions.**

Before leaving this section, we would like to comment on the nature of the problem. If we had data on (i) ball properties, (ii) batter approach properties, (iii) other properties (e.g. bats, weather, etc.), for each pitch, in addition to our pitching data –we could then attempt to decompose the proportion of variation of the different components and resolve the issue. The problem is we only have the pitch properties. With Statcast data, we have some batted ball results, which is a start for batter approach properties, but generally we lack the necessary data. Data could be summarized, in order to try to bring in variables, but such summarization likely masks the very variation we are trying to uncover. In short, definitive answers to this issue are difficult, if not impossible, to achieve.

3. What Changes in Pitching are Related to Home Run% in 2019?

Having established that quality of pitch is related to home runs, we want to determine what has changed in the pitching, if any, to understand the relationship between pitching changes and 2019's increase of home runs. In this section, we examine the six pitch components for changes. In (3.1) we discover the primary change is in *location,* in (3.2) we define the strike zone model, and in (3.3) we take a closer look at the specific changes in *location*.

---

each of the models used to estimate R^2 captures an aspect of the phenomenon. Averaging competing models produces a more accurate model.

[14] Wilson, Jason; Jordan Wong, Jeremiah Chuang, Wayne Greiner. [Explaining the MLB Home Run Record of 2017 with QOP](). Technical Report. 2018.

[15] Recognizing that 2019 is partial.

[16] Comparing the pseudo-R^2's in Appendices B and C, many are the same, and the largest difference was 0.01, most differences, where they exist, are 0.001. As a result, only the four seam and two seam fastballs (FF & FT) differ when averaged in Table 3.

3.1 Changes in Pitch Components

In our previous study[17], to identify pitching change we began by examining the trend of each pitch component by year. Control charts were used in order to determine whether observed changes were within historical range, or were extreme. The graphs in Figures 7 & 8 are called Control Charts and are routinely used in manufacturing quality control to detect when a process is within historic limits and when it is extreme. In particular, the middle lines are the mean of the component and the upper and lower limits (UCL and LCL) are the mean +/- three standard deviations. Thus, by viewing the graphs we can see the change in the component's behavior over the years, and when the change is particularly extreme (i.e. above the upper limit or below the lower limit). We show two control charts for each pitch type in Figures 7 & 8, with an additional type in Appendix A. On the left side are the control charts for all pitch types. On the right side are control charts for only pitches hit for home runs. Appendix A has the control charts for only pitches hit. The pattern for all pitches and home runs is about the same throughout, except for *location* in 2017 to 2019. Although the patterns are basically the same, the center and spread differs on the graphs by the set of pitches. The similarity of the pattern indicates that, in the aggregate, the batter hitting success, and home run production occurs on the same relative pitch properties encountered. The difference shows, in the aggregate, which pitch properties are more hittable and home run friendly. For example, for all pitches the mean vertical break is about 3.7 ft. with a range of 3.6 to 3.8, whereas for pitches hit, the mean is 3.63 with a range of 3.5 to 3.7, and for home runs it is 3.5 with a range of 3.4 to 3.6. Thus, the lower the vertical break, the easier the pitches are to hit well, on average.

In Figures 7 & 8, there are only two components which substantially changed from previous years: showing a decrease in horizontal break and an increase in location. The explanation for horizontal break will be given in the discussion of location, below.

Our scale for location is the only one of the six pitch components not on a physical scale based on direct measurement. It is a non-linear function of the (x,z) components of the ball's location in the strike zone plane. The scale starts with zero at the corners of the strike zone, and points accrue as the ball moves further away from the corners, both into the center of the strike zone as well as out of zone. The median location jumped from around 1.44 in 2018 to 1.48 in 2019, which is well above the historical norm. But how, specifically, did *location* quality degrade? The answer to this question includes a decrease in horizontal break, and more….

[17] Wilson, Jason; Jordan Wong, Jeremiah Chuang, Wayne Greiner. Explaining the MLB Home Run Record of 2017 with QOP. Technical Report. 2018.

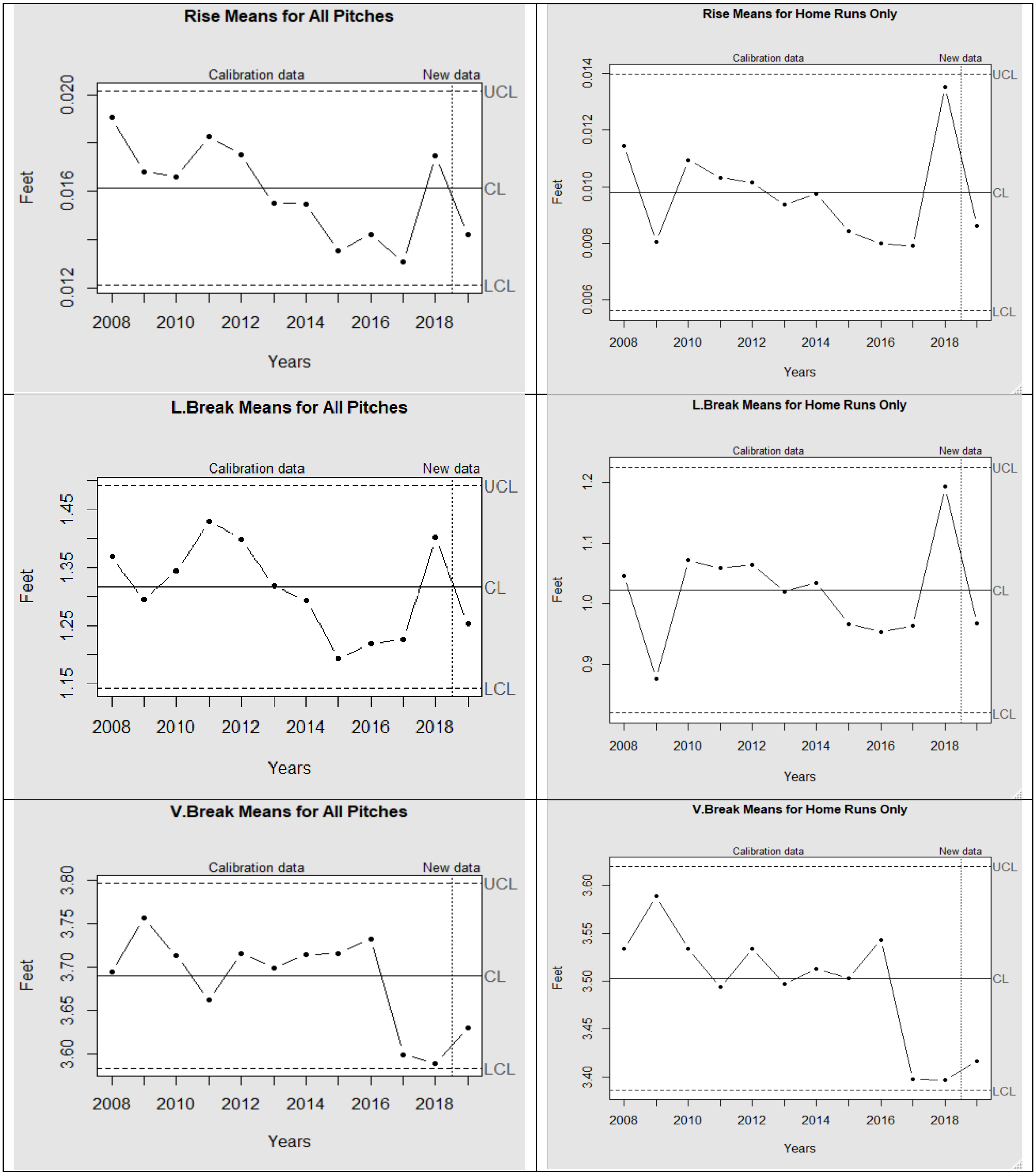


Figure 7. Control Charts for Rise, L.Break, and V.Break, for All Pitches and Home Runs Only

Figure 8. Control Charts for H.Break, Location, and Start Speed, for All Pitches and Home Runs Only

3.2 The Strike Zone Model

To examine location more closely, we used the PITCHf/x / Statcast location model, show below in Figure 9[18]. Zones 1-9 are inside the strike zone and zones 11-14 are outside.

The variable *zone* is provided in the PITCHf/x data from 2008 to 2018. However, in 2019 the variable is included but the cells are empty. Statcast also has a *zone* variable. However, the *zone* variable in Statcast does not match the PITCHf/x zone variable (Appendix F), probably because of how balls are binned which fall on the edges. Perhaps PITCHf/x considers balls on the edge of the strike zone to be in zones 11-14 whereas Statcast puts them in 1-4, 6-9. Because of these two problems, we developed our own algorithm, which counts border ties to be on the right side or top side of the border (Appendix F). Our counts turn out to be between the PITCHf/x and Statcast counts (Appendix F). For the purposes of the analysis below, all that really matters is a consistent zone measurement for every year, which we have constructed.

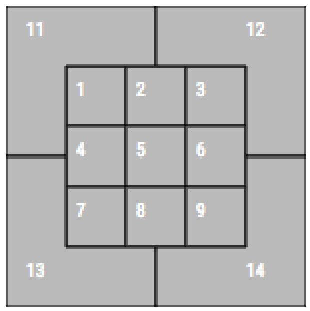


Figure 9. Strike zone location model.

3.3 How Location Changed

Zone alone, however, is not enough. In order to analyze location carefully, we must do two things: (1) determine what to measure and (2) observe the relevant distinctions. In this study, for (1) we chose to measure pitches in three different ways:

1. Distribution of pitches by zone: The count of pitches in each zone and divide them by the total number of pitches. The sum of all of the proportions is one.
2. Distribution of home runs by zone: The count of home runs in each zone and divide them by the total number of home runs. The sum of all of the proportions is one.
3. Proportion of home runs in each zone: The number of home runs in each zone divided by the total number of pitches to that zone. The sum of all proportions is not one.

In this study, for (2) we naturally chose to separate pitches by year. In addition, we separate pitches by pitcher-batter handedness match-up. This one distinction results in 13 zones x 4 match-ups x 3 measures = 156 graphs. This is extremely complicated. Additional distinctions could be made, such as by pitch type, which we did not do in this study, due to the added complexity, although we are considering it for further research. We draw three conclusions, below.

Overall, 2019 had fewer pitches in zones 4 & 5 along with more pitches in zone 14 (see Figures 10, 11, & 13). This represents a decrease in horizontal break and pitching lower in the strike zone. Other observations may be made from the Figures, particularly for 2015 to 2017, the years of documented changes in the baseball. We limit our conclusion, however, to changes in zones 4, 5, and 14 since they are objectively determined using the control charts. For a comparison of all zone changes together, and the specific numbers, see Figure 10 and Tables 5-7.

---

[18] Taken from Baseball Savant's documentation, https://baseballsavant.mlb.com/csv-docs.

The primary home run change in 2019 occurs from an increase in the proportion of home runs hit in every zone except zone 11 (Figure 10 and Table 7). When the splits are considered, the proportion of zones with an increase in home run proportions is 43/52 = 83% (Table 5).

	1	2	3	4	5	6	7	8	9	11	12	13	14
R-R	—	—	+	—	+	+	+	+	+	—	—	—	+
R-L	+	+	+	+	+	+	+	+	+	—	+	+	+
L-R	+	+	+	+	+	+	+	+	+	+	+	+	+
L-L	+	+	+	+	+	+	+	—	+	+	+	+	—

Table 5. Increases in home run proportion from 2018 to 2019? Columns are zones, rows are handedness splits, and cell entries are whether or not there was an increase in proportion of home runs from 2018 to 2019. For example, in zone 1 there was a decrease in home runs for R-R match-ups, but an increase for all other match-ups.

For handedness match-ups, there are several general observations, as seen in Appendix G. First, the highest proportion of pitches are out of the strike zone. Second, of the highest pitching zone proportion pitches, in the last few years pitchers have been increasing the proportion of pitches thrown low and close to the batter. For R-R and L-L, this is an increase in zones 14 and 13, respectively. These are the lowest proportion of home run zones (see Figure 10). For R-L and L-R, they have additionally decreased pitches to zones 11 & 12 (Appendix G). Third, of the inside the strike zone pitches, the balls have been slightly shifting to low and inside.

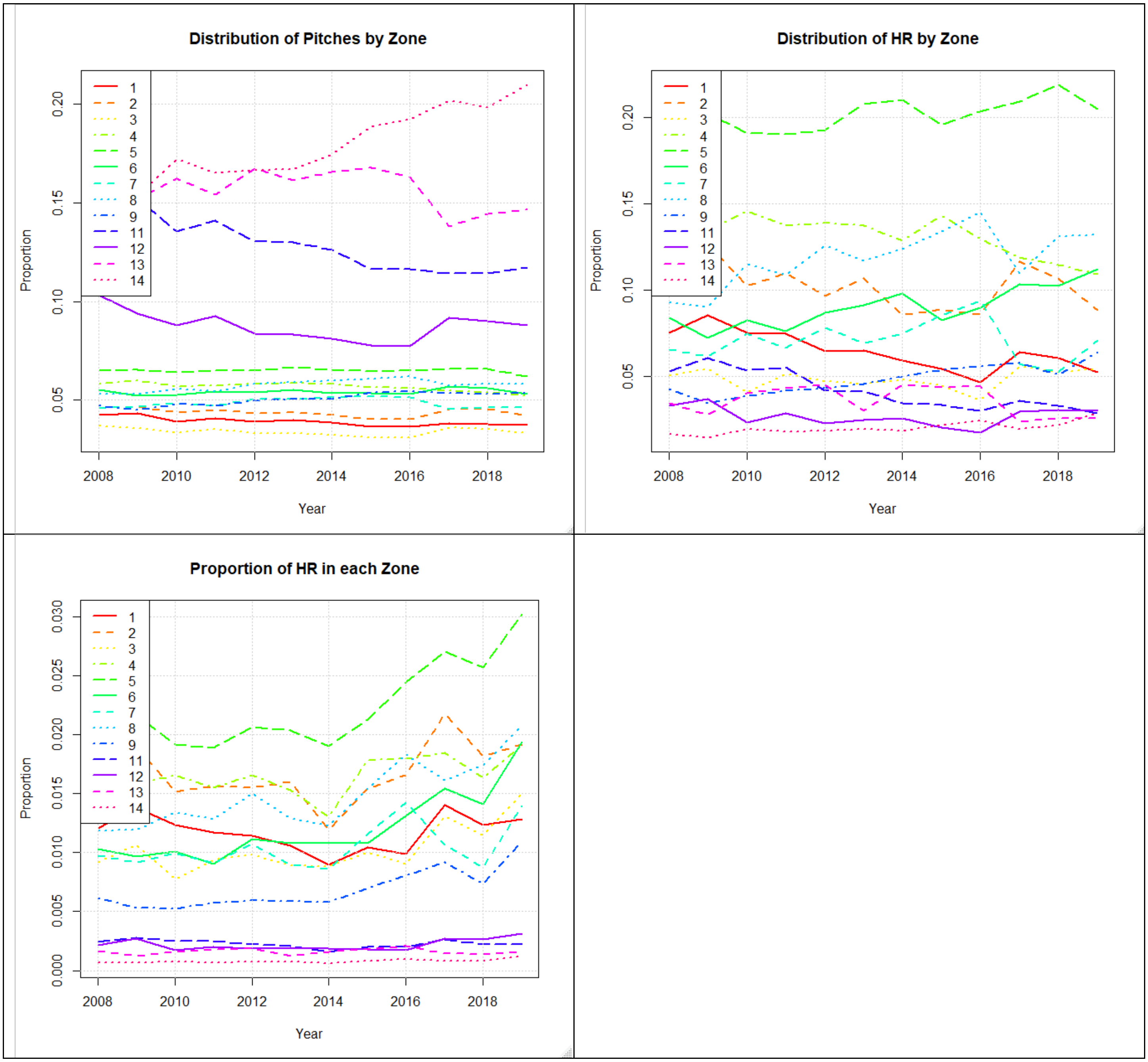


Figure 2. Comparison of the change in proportion of pitches in all zones, for all pitches

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.042	0.046	0.037	0.058	0.065	0.055	0.046	0.053	0.047	0.146	0.103	0.141	0.159
2009	0.043	0.047	0.036	0.060	0.066	0.052	0.047	0.053	0.045	0.152	0.094	0.151	0.154
2010	0.039	0.044	0.034	0.057	0.064	0.053	0.048	0.056	0.048	0.135	0.088	0.162	0.172
2011	0.041	0.045	0.035	0.057	0.065	0.054	0.047	0.055	0.047	0.141	0.092	0.154	0.165
2012	0.039	0.043	0.033	0.058	0.065	0.054	0.050	0.058	0.050	0.131	0.084	0.167	0.167
2013	0.040	0.044	0.033	0.059	0.067	0.055	0.050	0.059	0.051	0.130	0.083	0.162	0.167
2014	0.039	0.043	0.032	0.058	0.065	0.053	0.051	0.060	0.051	0.126	0.081	0.165	0.175
2015	0.036	0.040	0.031	0.056	0.065	0.053	0.052	0.061	0.054	0.116	0.078	0.168	0.189
2016	0.037	0.041	0.031	0.056	0.065	0.053	0.051	0.062	0.054	0.116	0.077	0.163	0.193
2017	0.038	0.046	0.036	0.055	0.066	0.057	0.045	0.058	0.054	0.114	0.092	0.138	0.202
2018	0.038	0.045	0.035	0.054	0.066	0.056	0.046	0.058	0.053	0.114	0.090	0.145	0.198
2019	0.037	0.042	0.033	0.052	0.062	0.053	0.046	0.058	0.053	0.117	0.088	0.147	0.210

Table 6. Table of distribution of pitches, by zone. E.g. In 2008, 0.042 = 4.2% of all pitches were in Zone 1

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.075	0.119	0.050	0.128	0.207	0.084	0.065	0.093	0.043	0.053	0.033	0.034	0.016
2009	0.085	0.127	0.054	0.135	0.202	0.072	0.061	0.090	0.034	0.060	0.037	0.027	0.014
2010	0.075	0.102	0.040	0.146	0.191	0.082	0.075	0.115	0.039	0.053	0.023	0.040	0.019
2011	0.075	0.110	0.051	0.137	0.190	0.076	0.066	0.109	0.042	0.055	0.028	0.043	0.018
2012	0.064	0.096	0.047	0.139	0.193	0.087	0.078	0.126	0.043	0.042	0.022	0.045	0.018
2013	0.065	0.107	0.045	0.138	0.208	0.091	0.069	0.117	0.045	0.041	0.025	0.030	0.019
2014	0.059	0.085	0.048	0.129	0.210	0.098	0.075	0.124	0.050	0.034	0.025	0.045	0.018
2015	0.054	0.088	0.045	0.143	0.196	0.082	0.085	0.134	0.054	0.033	0.020	0.043	0.021
2016	0.046	0.086	0.036	0.129	0.204	0.089	0.094	0.145	0.056	0.030	0.017	0.044	0.024
2017	0.064	0.117	0.056	0.119	0.209	0.103	0.057	0.110	0.058	0.036	0.030	0.024	0.019
2018	0.061	0.107	0.052	0.114	0.219	0.102	0.052	0.131	0.051	0.033	0.030	0.025	0.022
2019	0.052	0.088	0.054	0.109	0.205	0.112	0.071	0.132	0.064	0.028	0.030	0.025	0.028

Table 7. Table of distribution of home runs, by zone. E.g. In 2008, 0.075 = 7.5% of all home runs were from Zone 1.

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.012	0.018	0.009	0.015	0.022	0.010	0.010	0.012	0.006	0.0025	0.0022	0.0016	0.0007
2009	0.014	0.019	0.011	0.016	0.022	0.010	0.009	0.012	0.005	0.0028	0.0027	0.0013	0.0007
2010	0.012	0.015	0.008	0.017	0.019	0.010	0.010	0.013	0.005	0.0025	0.0017	0.0016	0.0007
2011	0.012	0.016	0.009	0.015	0.019	0.009	0.009	0.013	0.006	0.0025	0.0020	0.0018	0.0007
2012	0.011	0.015	0.010	0.017	0.021	0.011	0.011	0.015	0.006	0.0022	0.0019	0.0019	0.0008
2013	0.011	0.016	0.009	0.015	0.020	0.011	0.009	0.013	0.006	0.0021	0.0019	0.0012	0.0007
2014	0.009	0.012	0.009	0.013	0.019	0.011	0.009	0.012	0.006	0.0016	0.0018	0.0016	0.0006
2015	0.010	0.015	0.010	0.018	0.021	0.011	0.012	0.015	0.007	0.0020	0.0018	0.0018	0.0008
2016	0.010	0.017	0.009	0.018	0.024	0.013	0.014	0.018	0.008	0.0020	0.0017	0.0021	0.0010
2017	0.014	0.022	0.013	0.018	0.027	0.015	0.011	0.016	0.009	0.0026	0.0027	0.0015	0.0008
2018	0.012	0.018	0.011	0.016	0.026	0.014	0.009	0.017	0.007	0.0022	0.0026	0.0014	0.0008
2019	0.013	0.019	0.015	0.019	0.030	0.019	0.014	0.021	0.011	0.0022	0.0031	0.0016	0.0012

Table 8. Table of proportion of home runs, by zone. E.g. Of all pitches in Zone 1 in 2008, 0.012 = 1.2% were home runs.

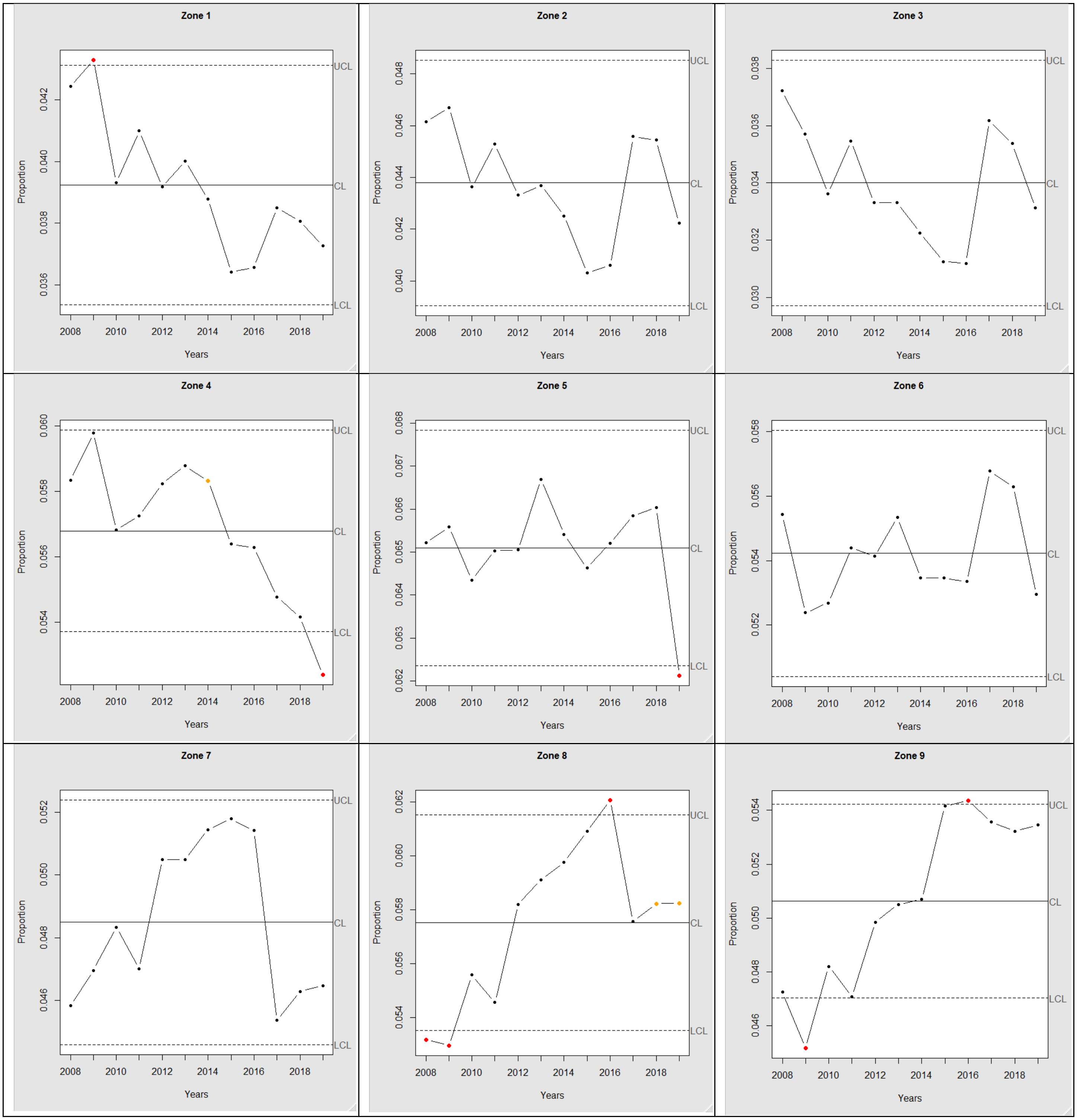


Figure 9. Control charts for zones 1-9, within the strike zone, for all pitches

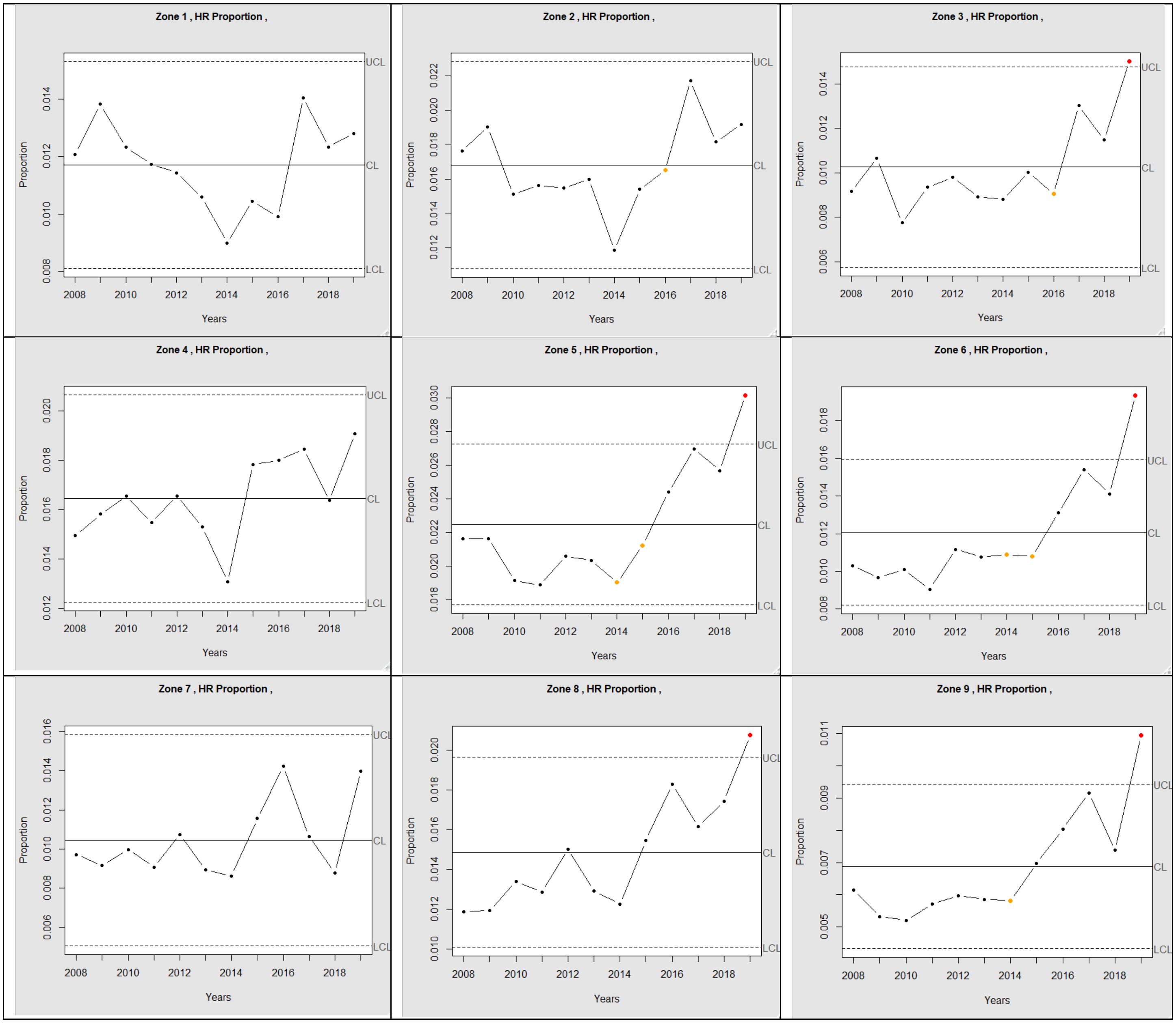


Figure 10. Control charts for zones 1-9, within the strike zone, for pitches resulting in home runs only

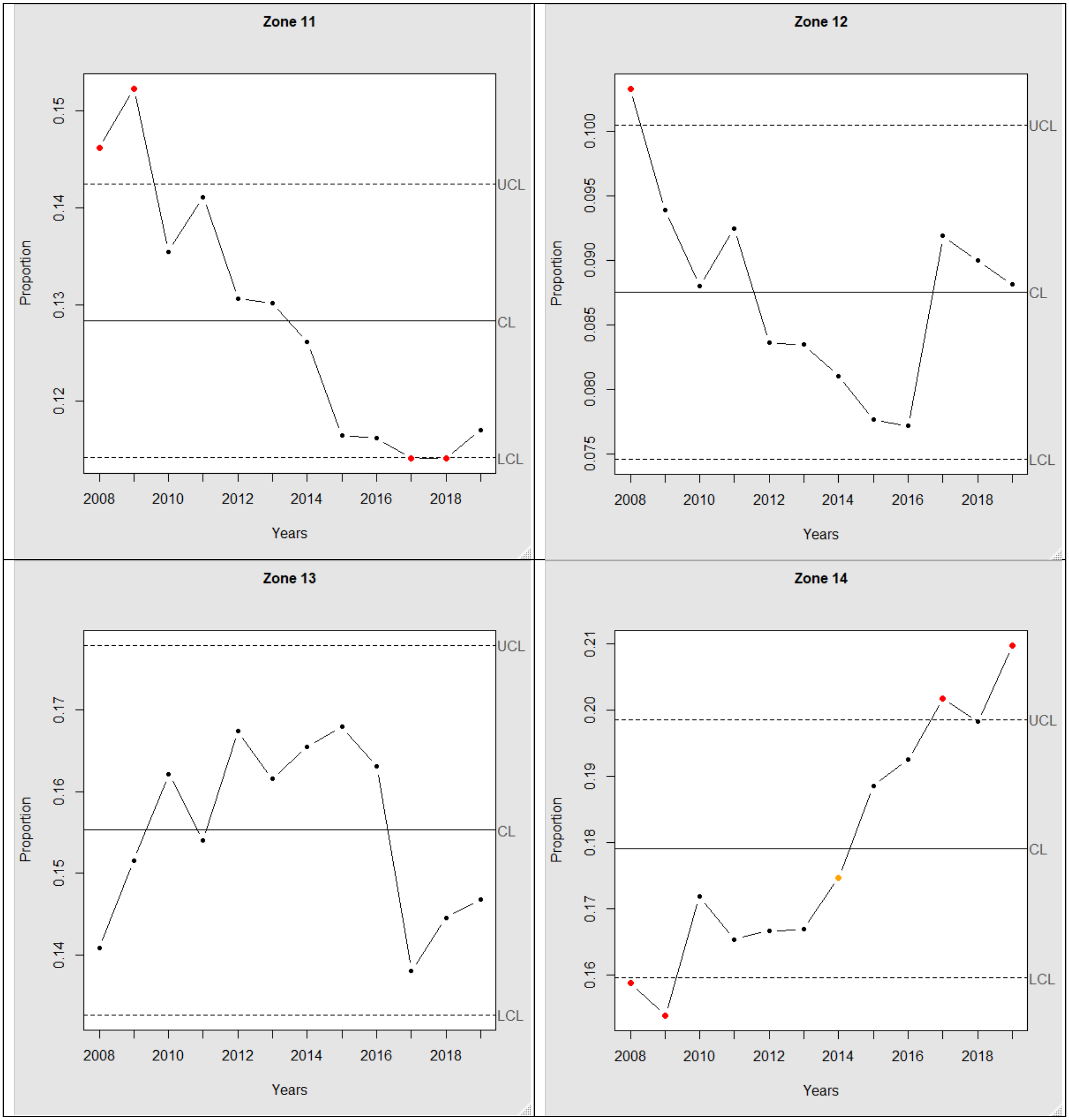


Figure 3. Control charts for zones 11-14, outside of the strike zone, for all pitches

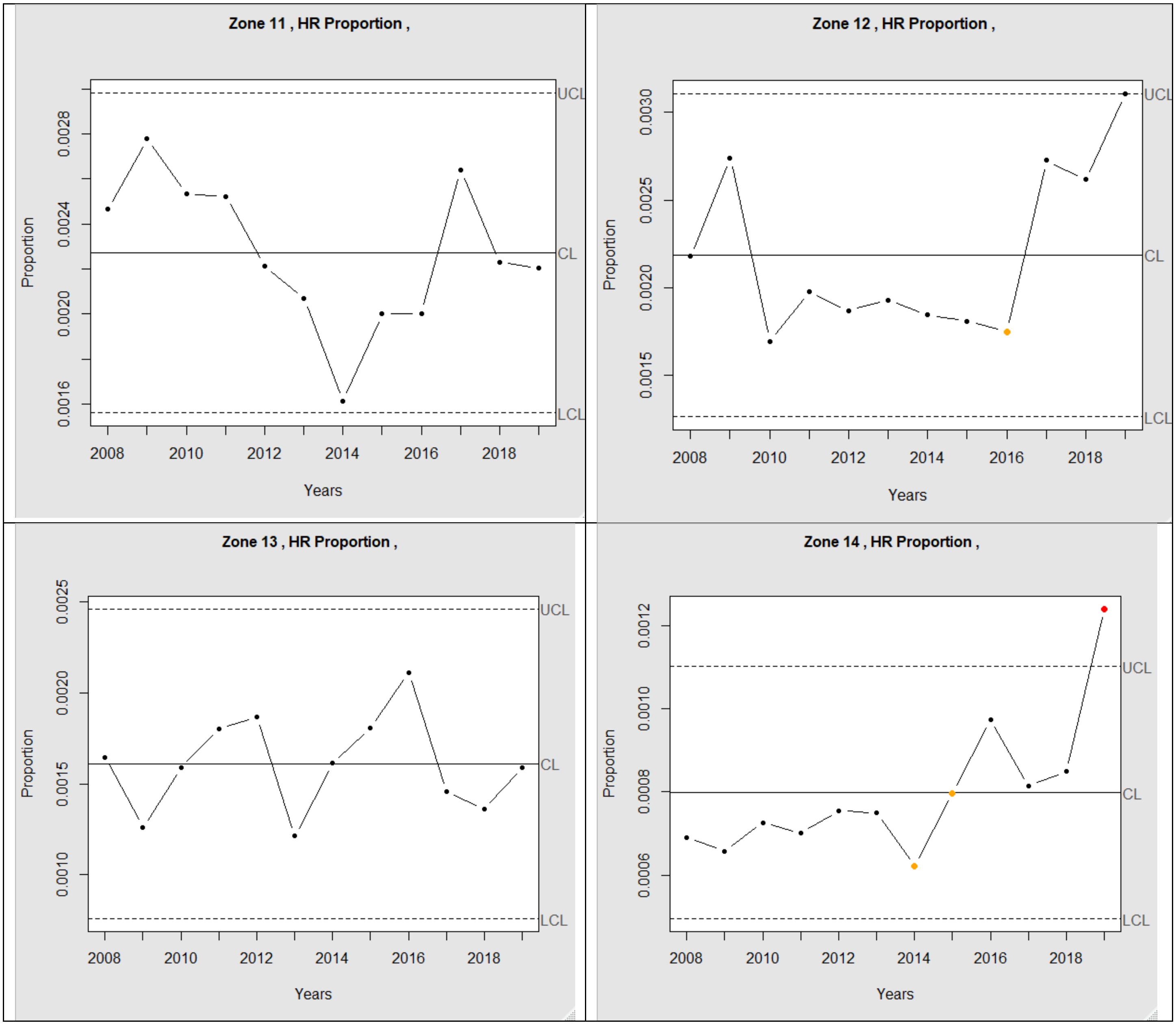


Figure 12. Control charts for zones 11-14, outside of the strike zone, for pitches resulting in home runs only

The proportion of home runs increased from 0.00085 in 2018 to 0.00124 in 2019. This is an increase of 0.00039 home runs per pitch. If this proportion held for the 2019 season, and the typical 720,000 pitches were thrown, and we recall 0.2098 (20.98%) of the pitches are in zone 14, this would result in an increase of 0.00039*720,000*0.2098 = 58.9 additional home runs. And that is just for zone 14. Similar calculations for each zone gives

zone	1	2	3	4	5	6	7	8	9	11	12	13	14
+HRs	12.6	30.9	84.2	102.1	200.4	200.0	173.6	140.8	137.7	-2.3	31.0	23.7	59.1

Summing the increased home runs for each zone is 1193.8, which is higher than the increase in home runs expected by commentators from 2018 to 2019. Note that these increases do **not** incorporate pitch quality. Further research into the relationship between pitch quality and these projections might be fruitful. To take such reasoning beyond speculation, the final numbers for the season would need to be obtained.

4. Conclusions

After establishing correlation between quality of pitch and home runs, we determined that quality of pitch accounts for a meaningful amount of the variation in the proportion of home runs in MLB. Two components substantially changed in 2019 from historic levels: *horizontal break* and *location*. Location had the proportionally largest change, and appears to be the primary factor. *location* significantly changed from the middle of the strike zone (zones 4 -6) to low and closer to the batter (zones 13 & 14, depending on batter handedness). This accounts for the decrease in horizontal break. At the same time, 83% of pitch zones experienced an increase in the proportion of home runs from 2018 to 2019 (counting differences in pitcher-batter handedness match-ups). These results are consistent with either a passive or an active pitcher, i.e. an unconscious reaction or conscious decision. The passive pitchers would be reacting to a perceived threat of batters hitting more home runs. The active pitchers would be attempting to control the game by altering their strategy, albeit unsuccessfully. Furthermore, balls with less drag may result in pitchers achieving less command and horizontal break: inadvertently pitching straighter. This may allow batters to better read the pitch trajectory and result in better contact.

Regardless of whether the change in pitching is unconscious, or not, we see that the quality of pitch in 2019 is projected to finish at a record low with home runs at a record high. Pitches are moving from locations that yield more home runs (middle of the strike zone) to locations that yield less home runs for batters (low and close to the batter). In the aggregate, this change is yielding the opposite of what may be expected (a decrease instead of an increase in home runs), in part due to lower quality pitches.

It must be kept in mind that the pitch quality variation accounts for around 26% to 40% of the variation in home run proportion, leaving 60% to 74% of the variation due to other factors. These other factors likely include changes in the ball, increased uppercut swinging by batters, and perhaps lesser factors as well. Therefore, pitch quality has been shown to be one of the factors in the home run surge of 2019, although not the majority factor.

Appendix A: Coefficients and p-values from General Linear Models by year

There are twelve sections of output below: the first six contain the regression coefficients for the generalized linear regression models, by year. The second six contain the corresponding p-values for each coefficient. Each section of output below consists of two tables. Both groups contain the pitch types, in order: CH, CU, FF, FT, SI, & SL. The top table is the coefficients, by year. The bottom table consists of three rows. The first row contains the means of the coefficients across years, the second row contains the standard deviations, and the third row contains the coefficient of variations.

```
> CHglm = pull.coeffs(HomeRun.GLMs, 1, 1)
             2008  2009  2010  2011  2012  2013  2014  2015  2016  2017  2018
(Intercept)  1.34  1.40  2.79  4.00  2.34  3.86  1.15  1.30  3.22  3.96  1.95
rise         0.79  4.99  6.88  5.17  8.85  3.96  4.95  7.91  4.61  3.76  5.69
breakpt     -0.19 -0.29 -0.30 -0.30 -0.35 -0.23 -0.35 -0.37 -0.35 -0.24 -0.30
tot.brk     -0.73 -0.70 -0.76 -0.72 -0.82 -0.81 -0.85 -0.79 -0.77 -0.81 -0.81
h.brk2      -0.21 -0.45 -0.26 -0.40 -0.61 -0.43 -0.41 -0.32 -0.46 -0.32 -0.74
loc         -0.04 -0.05 -0.10 -0.04 -0.07 -0.02 -0.02 -0.08 -0.08 -0.02 -0.04
MPH         -0.04 -0.03 -0.05 -0.07 -0.04 -0.06 -0.03 -0.03 -0.05 -0.06 -0.03
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       2.48        5.23       -0.30       -0.78       -0.42       -0.05       -0.04
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       1.14        2.17        0.06        0.05        0.15        0.03        0.01
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       2.18        2.41       -5.22      -16.32       -2.74       -1.75       -3.10
> CUglm = pull.coeffs(HomeRun.GLMs, 2, 1)
             2008  2009  2010  2011  2012  2013  2014  2015  2016  2017  2018
(Intercept) 10.23  9.28  9.82  3.40  4.52  6.02  5.78  9.16  6.36  8.38  4.28
rise        -2.03 -3.26 -4.31 -3.74 -6.13 -2.06 -5.21 -5.43 -7.22 -4.91 -4.13
breakpt     -0.08 -0.11 -0.06 -0.02  0.02 -0.07  0.01 -0.01  0.03 -0.04 -0.01
tot.brk     -0.92 -0.94 -0.98 -0.83 -1.00 -1.00 -0.75 -0.93 -0.74 -0.93 -0.83
h.brk2      -0.33 -0.14 -0.15  0.09 -0.12 -0.24 -0.32 -0.25 -0.27 -0.54 -0.30
loc         -0.05 -0.14 -0.04 -0.05  0.01  0.00 -0.10 -0.08 -0.06  0.01 -0.02
MPH         -0.14 -0.13 -0.13 -0.07 -0.07 -0.09 -0.09 -0.13 -0.10 -0.11 -0.07
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       7.02       -4.40       -0.03       -0.90       -0.23       -0.05       -0.10
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       2.44        1.61        0.04        0.09        0.16        0.05        0.03
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       2.88       -2.74       -0.67       -9.63       -1.47       -0.99       -3.66
> FFglm = pull.coeffs(HomeRun.GLMs, 3, 1)
             2008  2009  2010  2011  2012  2013  2014  2015  2016  2017  2018
(Intercept)  1.19  0.84  0.24  1.86  1.69  1.83  2.38  1.27  2.71  2.05  3.02
rise         0.93  3.40  8.01  6.06  3.27 -1.65  6.47  6.32  6.60  7.84  2.54
breakpt     -0.14 -0.16 -0.23 -0.22 -0.20 -0.15 -0.19 -0.22 -0.20 -0.24 -0.13
tot.brk     -0.13 -0.14 -0.13 -0.12 -0.13 -0.15 -0.12 -0.14 -0.08 -0.09 -0.04
h.brk2       0.09  0.16  0.12 -0.06  0.14  0.04  0.10  0.04  0.13  0.16  0.23
loc         -0.04 -0.07 -0.05 -0.06 -0.05 -0.01 -0.03 -0.03 -0.03 -0.01  0.02
MPH         -0.06 -0.06 -0.05 -0.07 -0.07 -0.07 -0.07 -0.06 -0.08 -0.07 -0.08
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       1.73        4.53       -0.19       -0.12        0.10       -0.03       -0.07
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.82        3.08        0.04        0.03        0.08        0.03        0.01
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       2.12        1.47       -4.94       -3.48        1.38       -1.20       -6.93
FTglm = pull.coeffs(HomeRun.GLMs, 4, 1)
             2008  2009   2010    2011  2012  2013  2014  2015  2016  2017  2018
(Intercept)  0.24  1.51   1.96    3.91  2.09  2.33  4.30 -0.13  0.29  2.61  3.54
rise         2.80  4.77 -65.41 -229.85  6.58 -9.31  5.99  4.60 12.39 12.90 13.57
breakpt     -0.16 -0.24  -0.15    0.06 -0.27 -0.13 -0.24 -0.22 -0.38 -0.31 -0.30
tot.brk     -0.22 -0.30  -0.32   -0.39 -0.31 -0.38 -0.41 -0.33 -0.22 -0.35 -0.23
h.brk2      -0.25 -0.05  -0.22   -0.29 -0.62 -0.70 -0.66 -0.24 -0.24 -0.09 -0.55
loc         -0.07 -0.06  -0.10   -0.02 -0.06 -0.01  0.03 -0.05 -0.05  0.00 -0.02
MPH         -0.04 -0.06  -0.06   -0.08 -0.06 -0.06 -0.08 -0.04 -0.04 -0.06 -0.07
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       2.06      -21.91       -0.21       -0.31       -0.36       -0.04       -0.06
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
```

```
       1.50        72.49         0.12         0.07         0.24         0.04         0.01
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       1.37        -0.30        -1.80        -4.64        -1.51        -0.98        -3.93
> SIglm = pull.coeffs(HomeRun.GLMs, 5, 1)
             2008  2009  2010  2011  2012  2013  2014  2015  2016  2017  2018
(Intercept)  0.91  1.62  1.75  1.06  2.53  3.52  2.43  3.81  1.27  3.15  0.57
rise         0.47 -0.14  1.88  0.34  0.87  1.24  0.88 -0.77 -1.67  0.74  0.65
breakpt     -0.08 -0.09 -0.12 -0.08 -0.10 -0.10 -0.08 -0.08 -0.01 -0.10 -0.10
tot.brk     -0.31 -0.34 -0.36 -0.27 -0.35 -0.40 -0.27 -0.35 -0.32 -0.25 -0.32
h.brk2      -0.23 -0.49 -0.29 -0.69 -0.38 -0.56 -0.63 -0.45 -0.49 -0.37 -0.40
loc         -0.04 -0.05 -0.06 -0.09 -0.07 -0.01 -0.09 -0.07 -0.01  0.00  0.01
MPH         -0.05 -0.05 -0.06 -0.05 -0.06 -0.07 -0.06 -0.08 -0.05 -0.07 -0.04
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       2.06         0.41        -0.08        -0.32        -0.45        -0.04        -0.06
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       1.10         0.97         0.03         0.05         0.14         0.04         0.01
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       1.86         0.42        -3.00        -7.18        -3.29        -1.19        -5.00
> SLglm = pull.coeffs(HomeRun.GLMs, 6, 1)
             2008  2009  2010  2011  2012  2013  2014  2015  2016  2017  2018
(Intercept)  3.61  3.59  4.80  5.05  3.06  6.22  5.19  5.50  4.99  3.94  3.60
rise         2.96  3.02  0.10  1.69  1.41  3.07 -1.31 -1.02  1.42  3.01  1.58
breakpt     -0.20 -0.20 -0.15 -0.19 -0.15 -0.21 -0.16 -0.14 -0.19 -0.22 -0.15
tot.brk     -0.74 -0.69 -0.77 -0.89 -0.68 -0.75 -0.78 -0.79 -0.78 -0.76 -0.81
h.brk2      -0.23 -0.32 -0.33 -0.18 -0.19 -0.27 -0.13 -0.20 -0.15 -0.28 -0.41
loc          0.05  0.04  0.00  0.00 -0.01  0.03  0.03  0.00  0.01  0.03  0.06
MPH         -0.07 -0.07 -0.08 -0.08 -0.06 -0.10 -0.08 -0.09 -0.08 -0.07 -0.06
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       4.50         1.45        -0.18        -0.77        -0.24         0.02        -0.07
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       0.99         1.60         0.03         0.06         0.09         0.02         0.01
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       4.54         0.91        -5.98       -13.52        -2.81         0.91        -6.67
> CHglm = pull.coeffs(HomeRun.GLMs, 1, 4)
            2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
(Intercept) 0.22 0.21 0.01 0.00 0.04 0.00 0.34 0.23 0.00 0.00 0.09
rise        0.71 0.00 0.00 0.00 0.00 0.14 0.04 0.00 0.05 0.00 0.00
breakpt     0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
tot.brk     0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
h.brk2      0.05 0.00 0.02 0.00 0.00 0.00 0.00 0.01 0.00 0.01 0.00
loc         0.23 0.14 0.00 0.21 0.02 0.47 0.65 0.01 0.01 0.60 0.22
MPH         0.00 0.01 0.00 0.00 0.00 0.00 0.06 0.02 0.00 0.00 0.01
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       0.10         0.09         0.00         0.00         0.01         0.23         0.01
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       0.12         0.21         0.00         0.00         0.02         0.24         0.02
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       0.85         0.41         0.36         0.30         0.49         0.97         0.58
> CUglm = pull.coeffs(HomeRun.GLMs, 2, 4)
            2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
(Intercept) 0.00 0.00 0.00 0.12 0.02 0.00 0.01 0.00 0.00 0.00 0.01
rise        0.06 0.03 0.00 0.01 0.00 0.12 0.00 0.00 0.00 0.00 0.00
breakpt     0.04 0.01 0.16 0.66 0.61 0.09 0.87 0.86 0.41 0.29 0.77
tot.brk     0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
h.brk2      0.02 0.39 0.30 0.57 0.40 0.10 0.03 0.08 0.04 0.00 0.03
loc         0.22 0.00 0.29 0.23 0.75 0.97 0.03 0.04 0.11 0.71 0.54
MPH         0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       0.01         0.02         0.43         0.00         0.18         0.35         0.00
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       0.04         0.04         0.33         0.00         0.20         0.34         0.00
(Intercept)         rise      breakpt      tot.brk       h.brk2          loc          MPH
       0.40         0.55         1.29         0.44         0.90         1.05         0.35
> FFglm = pull.coeffs(HomeRun.GLMs, 3, 4)
            2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
(Intercept) 0.08 0.23 0.76 0.02 0.03 0.03 0.01 0.10 0.00 0.00 0.00
rise        0.70 0.01 0.00 0.01 0.42 0.85 0.02 0.06 0.08 0.00 0.00
breakpt     0.01 0.00 0.00 0.00 0.00 0.06 0.00 0.00 0.00 0.00 0.00
```

```
tot.brk      0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.11
h.brk2       0.26 0.04 0.14 0.50 0.07 0.60 0.22 0.57 0.03 0.01 0.00
loc          0.04 0.00 0.01 0.00 0.01 0.46 0.16 0.09 0.10 0.50 0.11
MPH          0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.11        0.20        0.01        0.01        0.22        0.14        0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.22        0.31        0.02        0.03        0.23        0.18        0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.51        0.63        0.38        0.31        0.96        0.76        0.30
FTglm = pull.coeffs(HomeRun.GLMs, 4, 4)
            2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
(Intercept) 0.87 0.29 0.21 0.04 0.16 0.10 0.00 0.92 0.84 0.04 0.01
rise        0.38 0.17 0.43 0.19 0.23 0.64 0.16 0.67 0.04 0.26 0.17
breakpt     0.06 0.03 0.57 0.86 0.01 0.33 0.00 0.04 0.00 0.00 0.01
tot.brk     0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
h.brk2      0.10 0.77 0.17 0.06 0.00 0.00 0.00 0.10 0.09 0.45 0.00
loc         0.04 0.04 0.00 0.62 0.05 0.82 0.40 0.09 0.07 0.92 0.57
MPH         0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.01 0.00 0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.32        0.30        0.17        0.00        0.16        0.33        0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.37        0.20        0.29        0.00        0.24        0.35        0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.86        1.49        0.59        0.38        0.66        0.94        0.57
> SIglm = pull.coeffs(HomeRun.GLMs, 5, 4)
            2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
(Intercept) 0.44 0.20 0.19 0.43 0.09 0.07 0.24 0.03 0.42 0.08 0.74
rise        0.42 0.91 0.12 0.80 0.61 0.59 0.39 0.72 0.23 0.61 0.56
breakpt     0.00 0.01 0.00 0.03 0.01 0.03 0.03 0.08 0.75 0.01 0.01
tot.brk     0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
h.brk2      0.08 0.00 0.05 0.00 0.02 0.00 0.00 0.02 0.00 0.04 0.07
loc         0.18 0.09 0.06 0.00 0.05 0.83 0.02 0.08 0.72 0.94 0.73
MPH         0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.02
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.27        0.54        0.09        0.00        0.03        0.34        0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.22        0.24        0.22        0.00        0.03        0.38        0.01
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       1.23        2.29        0.40        0.46        0.87        0.89        0.46
> SLglm = pull.coeffs(HomeRun.GLMs, 6, 4)
            2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
(Intercept) 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.00
rise        0.04 0.06 0.96 0.30 0.28 0.08 0.62 0.66 0.43 0.12 0.03
breakpt     0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
tot.brk     0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
h.brk2      0.05 0.01 0.01 0.14 0.08 0.02 0.34 0.11 0.16 0.01 0.00
loc         0.07 0.11 0.95 0.99 0.59 0.20 0.31 1.00 0.55 0.23 0.01
MPH         0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.00        0.33        0.00        0.00        0.08        0.46        0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.00        0.31        0.00        0.00        0.10        0.38        0.00
(Intercept)        rise     breakpt     tot.brk      h.brk2         loc         MPH
       0.46        1.05        0.42        0.30        0.81        1.20        0.36
```

Appendix B: Generalized Linear Models, by pitch type, with year as a factor

In this appendix we show a general linear logistic regression model, of the form shown in the function call. The variable "year" was entered as a factor, in order to see whether there were effects due to particular years. Six models are shown, one for each pitch type. The striking feature is the small p-values of the pitch components, as well as the consistent 2016, 2017, and 2019 years as varying from the rest. The signs of the coefficients are as expected, except for *location*, which is negative, meaning that poorer location (since zero is the best location score) results in decreased home runs. This is because over half of the pitches are not swung at, many due to being out of the strike zone. When only pitches swung at are considered, the location coefficients become positive, as expected. In this Appendix the 'all pitches' model is shown first, followed by the 'pitches swung at only'.

Appendix B.1. All pitches

```
> for (i in 1:6) run.GLM(All.pit, pitch.type=pitch.types[i])
[1] "***** OUTPUT FOR CH *****"


Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 + 
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),])


Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-1.1871  -0.1437  -0.1156  -0.0863   5.5019  


Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)  2.144895   0.322057   6.660 2.74e-11 ***
year2009     0.048228   0.061825   0.780 0.435351    
year2010    -0.047488   0.061287  -0.775 0.438435    
year2011    -0.044671   0.062183  -0.718 0.472522    
year2012     0.124787   0.060332   2.068 0.038607 *  
year2013     0.016657   0.061766   0.270 0.787411    
year2014    -0.041392   0.063095  -0.656 0.511806    
year2015     0.072214   0.061026   1.183 0.236677    
year2016     0.201980   0.059889   3.373 0.000745 ***
year2017     0.215491   0.059768   3.605 0.000312 ***
year2018     0.093787   0.060875   1.541 0.123402    
year2019     0.282285   0.066293   4.258 2.06e-05 ***
rise         3.870810   0.282583  13.698  < 2e-16 ***
breakpt     -0.260079   0.011026 -23.588  < 2e-16 ***
tot.brk     -0.722091   0.017627 -40.964  < 2e-16 ***
h.brk2      -0.415524   0.033874 -12.267  < 2e-16 ***
loc         -0.130281   0.008018 -16.249  < 2e-16 ***
MPH         -0.042619   0.003738 -11.402  < 2e-16 ***
---
(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 77267  on 855627  degrees of freedom
Residual deviance: 74300  on 855610  degrees of freedom
AIC: 74336


Number of Fisher Scoring iterations: 8
```

```
[1] "Predicted HRs:  6590 Actual HRs: 6592"
       McFadden      McFaddenAdj         CoxSnell       Nagelkerke   AldrichNelson VeallZimmermann           Effron
          0.895            0.895            0.524            0.930           0.184           0.914            0.002
McKelveyZavoina             Tjur              AIC              BIC          logLik         logLik0               G2
          0.173            0.003        74336.026        74545.899      -37150.013     -354451.146       634602.266


[1] "***** OUTPUT FOR CU *****"


Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),])


Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3975  -0.1192  -0.0977  -0.0744   3.9436


Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.773490   0.540450  12.533  < 2e-16 ***
year2009     0.020777   0.084681   0.245 0.806179
year2010     0.125662   0.082122   1.530 0.125971
year2011    -0.039484   0.086867  -0.455 0.649443
year2012     0.113126   0.081656   1.385 0.165931
year2013    -0.014760   0.084440  -0.175 0.861236
year2014     0.045433   0.085431   0.532 0.594861
year2015     0.216325   0.083437   2.593 0.009523 **
year2016     0.263369   0.079638   3.307 0.000943 ***
year2017     0.274247   0.078795   3.481 0.000500 ***
year2018     0.256556   0.079156   3.241 0.001190 **
year2019     0.286202   0.088472   3.235 0.001217 **
rise        -4.369090   0.405661 -10.770  < 2e-16 ***
breakpt     -0.028716   0.011706  -2.453 0.014159 *
tot.brk     -0.886741   0.035302 -25.119  < 2e-16 ***
h.brk2      -0.254864   0.041574  -6.130 8.77e-10 ***
loc         -0.105016   0.010205 -10.290  < 2e-16 ***
MPH         -0.099708   0.005582 -17.863  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 46499  on 697807  degrees of freedom
Residual deviance: 44857  on 697790  degrees of freedom
AIC: 44893


Number of Fisher Scoring iterations: 9


[1] "Predicted HRs:  3733 Actual HRs: 3734"
       McFadden      McFaddenAdj         CoxSnell       Nagelkerke   AldrichNelson VeallZimmermann           Effron
          0.937            0.937            0.614            0.962           0.224           0.951            0.002
McKelveyZavoina             Tjur              AIC              BIC          logLik         logLik0               G2
          0.194            0.002        44893.034        45099.236      -22428.517     -354451.146       664045.259
```

```
[1] "***** OUTPUT FOR FF *****"


Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])


Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7697  -0.1363  -0.1246  -0.1133   3.9917


Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.354761   0.212709   6.369 1.90e-10 ***
year2009     0.087997   0.033204   2.650 0.008045 **
year2010    -0.054500   0.034399  -1.584 0.113117
year2011    -0.024495   0.034424  -0.712 0.476743
year2012     0.098589   0.033211   2.969 0.002992 **
year2013     0.004698   0.033784   0.139 0.889406
year2014    -0.075226   0.034897  -2.156 0.031108 *
year2015     0.122888   0.033035   3.720 0.000199 ***
year2016     0.222458   0.032186   6.912 4.79e-12 ***
year2017     0.343259   0.031908  10.758  < 2e-16 ***
year2018     0.262630   0.032057   8.193 2.56e-16 ***
year2019     0.438711   0.034401  12.753  < 2e-16 ***
rise         3.725254   0.330728  11.264  < 2e-16 ***
breakpt     -0.183281   0.013136 -13.953  < 2e-16 ***
tot.brk     -0.048934   0.008001  -6.116 9.62e-10 ***
h.brk2       0.132680   0.021273   6.237 4.46e-10 ***
loc         -0.103804   0.004431 -23.429  < 2e-16 ***
MPH         -0.065186   0.002275 -28.651  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 267957  on 2900994  degrees of freedom
Residual deviance: 265803  on 2900977  degrees of freedom
AIC: 265839


Number of Fisher Scoring iterations: 9


[1] "Predicted HRs:  22961 Actual HRs: 22972"
      McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann          Effron
         0.625           0.625           0.142           0.653           0.044           0.642           0.001
McKelveyZavoina            Tjur             AIC             BIC          logLik         logLik0              G2
         0.043           0.001      265838.522      266070.372     -132901.261     -354451.146      443099.770


[1] "***** OUTPUT FOR FT *****"


Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
```

```
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])


Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-0.4745  -0.1289  -0.1158  -0.1033   4.2424  


Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)  1.701303   0.412825   4.121 3.77e-05 ***
year2009     0.090575   0.061496   1.473 0.140789    
year2010    -0.014156   0.061902  -0.229 0.819112    
year2011    -0.053312   0.063111  -0.845 0.398260    
year2012     0.006597   0.061378   0.107 0.914401    
year2013    -0.024136   0.060856  -0.397 0.691658    
year2014    -0.193378   0.063718  -3.035 0.002406 ** 
year2015     0.045145   0.060828   0.742 0.457989    
year2016     0.221428   0.060509   3.659 0.000253 ***
year2017     0.278465   0.058219   4.783 1.73e-06 ***
year2018     0.164719   0.060766   2.711 0.006714 ** 
year2019     0.298577   0.072151   4.138 3.50e-05 ***
rise         5.842168   1.157781   5.046 4.51e-07 ***
breakpt     -0.275881   0.026741 -10.317  < 2e-16 ***
tot.brk     -0.267212   0.015400 -17.352  < 2e-16 ***
h.brk2      -0.330787   0.043797  -7.553 4.26e-14 ***
loc         -0.093558   0.008067 -11.598  < 2e-16 ***
MPH         -0.056740   0.004426 -12.819  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 82123  on 994774  degrees of freedom
Residual deviance: 81260  on 994757  degrees of freedom
AIC: 81296


Number of Fisher Scoring iterations: 9


[1] "Predicted HRs:  6877 Actual HRs: 6880"
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann          Effron 
          0.885           0.885           0.468           0.918           0.161           0.904           0.001 
 McKelveyZavoina            Tjur             AIC             BIC          logLik         logLik0              G2 
          0.053           0.001       81295.587       81508.172      -40629.794     -354451.146      627642.705 


[1] "***** OUTPUT FOR SI *****"


Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 + 
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])


Deviance Residuals: 
```

```
    Min       1Q   Median       3Q      Max
-0.2646  -0.1276  -0.1138  -0.1006   3.5791


Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.610317   0.439576   3.663 0.000249 ***
year2009     0.076492   0.055584   1.376 0.168775
year2010    -0.015668   0.056024  -0.280 0.779740
year2011     0.035694   0.057353   0.622 0.533715
year2012     0.041283   0.059827   0.690 0.490173
year2013    -0.058467   0.065452  -0.893 0.371705
year2014    -0.177045   0.066696  -2.655 0.007943 **
year2015     0.024271   0.064411   0.377 0.706305
year2016     0.127134   0.060443   2.103 0.035433 *
year2017     0.250132   0.063625   3.931 8.45e-05 ***
year2018     0.089964   0.063113   1.425 0.154032
year2019     0.411301   0.069343   5.931 3.00e-09 ***
rise         0.481266   0.304681   1.580 0.114204
breakpt     -0.083128   0.009342  -8.898  < 2e-16 ***
tot.brk     -0.289201   0.017796 -16.251  < 2e-16 ***
h.brk2      -0.418483   0.048601  -8.611  < 2e-16 ***
loc         -0.099948   0.008914 -11.213  < 2e-16 ***
MPH         -0.054897   0.004724 -11.620  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 68139  on 842019  degrees of freedom
Residual deviance: 67413  on 842002  degrees of freedom
AIC: 67449


Number of Fisher Scoring iterations: 8


[1] "Predicted HRs:  5683 Actual HRs: 5685"
       McFadden     McFaddenAdj        CoxSnell       Nagelkerke   AldrichNelson VeallZimmermann          Effron
          0.905           0.905           0.533            0.937           0.188           0.923           0.001
McKelveyZavoina            Tjur             AIC              BIC          logLik         logLik0              G2
          0.042           0.001       67448.768        67658.352      -33706.384     -354451.146      641489.525


[1] "***** OUTPUT FOR SL *****"


Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])


Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9973  -0.1344  -0.1102  -0.0863   3.6969


Coefficients:
```

```
               Estimate Std. Error z value Pr(>|z|)    
(Intercept)  4.194098   0.332689  12.607  < 2e-16 ***
year2009     0.029543   0.054479   0.542   0.5876    
year2010    -0.021591   0.055428  -0.390   0.6969    
year2011     0.022359   0.054376   0.411   0.6809    
year2012     0.117335   0.053629   2.188   0.0287 *  
year2013     0.058100   0.054055   1.075   0.2824    
year2014     0.002513   0.056570   0.044   0.9646    
year2015     0.103358   0.055462   1.864   0.0624 .  
year2016     0.308747   0.052175   5.918 3.27e-09 ***
year2017     0.294771   0.050603   5.825 5.71e-09 ***
year2018     0.129643   0.051817   2.502   0.0124 *  
year2019     0.391155   0.054139   7.225 5.01e-13 ***
rise         1.577311   0.273269   5.772 7.83e-09 ***
breakpt     -0.175443   0.006790 -25.839  < 2e-16 ***
tot.brk     -0.739373   0.016957 -43.603  < 2e-16 ***
h.brk2      -0.281330   0.033491  -8.400  < 2e-16 ***
loc         -0.063999   0.006540  -9.786  < 2e-16 ***
MPH         -0.071947   0.003693 -19.481  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 105215  on 1290146  degrees of freedom
Residual deviance: 102052  on 1290129  degrees of freedom
AIC: 102088


Number of Fisher Scoring iterations: 8
```

```
[1] "Predicted HRs:  8789 Actual HRs: 8790"
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann          Effron 
          0.856           0.856           0.375           0.888           0.125           0.874           0.002 
McKelveyZavoina            Tjur             AIC             BIC          logLik         logLik0              G2 
          0.135           0.002      102088.195      102305.460      -51026.097     -354451.146      606850.098 
```

Appendix B.2. Only pitches swung at

These models are the same as above, with the exception that it is produced on a subset of only pitches that were swung at, whereas the models above were built from all pitches. The singular purpose for including this section is to address the change in the sign of the *loc* coefficient to positive, since they are negative in the sections above. See discussion in the body of the paper.

```
> for (i in 1:6) run.GLM(All.pit, pitch.type=pitch.types[i])
[1] "***** OUTPUT FOR CH *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7065  -0.1957  -0.1585  -0.1290   3.5388

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.200099   0.323143   6.808 9.87e-12 ***
year2009     0.063975   0.062154   1.029  0.30335
year2010    -0.064573   0.061583  -1.049  0.29438
year2011    -0.086189   0.062509  -1.379  0.16795
year2012     0.090804   0.060635   1.498  0.13425
year2013    -0.019864   0.062060  -0.320  0.74891
year2014    -0.091959   0.063383  -1.451  0.14682
year2015     0.013995   0.061368   0.228  0.81960
year2016     0.152330   0.060180   2.531  0.01137 *
year2017     0.169530   0.060113   2.820  0.00480 **
year2018     0.038451   0.061184   0.628  0.52971
year2019     0.212430   0.066641   3.188  0.00143 **
rise         2.104368   0.370336   5.682 1.33e-08 ***
breakpt     -0.180190   0.011884 -15.162  < 2e-16 ***
tot.brk     -0.819937   0.019541 -41.959  < 2e-16 ***
h.brk2      -0.432605   0.033931 -12.749  < 2e-16 ***
loc          0.116493   0.008937  13.035  < 2e-16 ***
MPH         -0.035666   0.003724  -9.577  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68422  on 438984  degrees of freedom
Residual deviance: 66263  on 438967  degrees of freedom
AIC: 66299

Number of Fisher Scoring iterations: 7

[1] "Predicted HRs:  6590 Actual HRs: 6592"
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann          Effron
          0.892           0.892           0.714           0.947           0.276           0.922           0.004
McKelveyZavoina            Tjur             AIC             BIC          logLik         logLik0              G2
          0.091           0.005       66299.351       66497.211      -33131.675     -308027.089      549790.828

[1] "***** OUTPUT FOR CU *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8762  -0.1842  -0.1506  -0.1207   3.4759

Coefficients:
```

```
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  7.370335   0.534392  13.792  < 2e-16 ***
year2009     0.031860   0.085137   0.374  0.70824    
year2010     0.105944   0.082601   1.283  0.19963    
year2011    -0.089556   0.087303  -1.026  0.30498    
year2012     0.091202   0.082145   1.110  0.26689    
year2013    -0.011268   0.084915  -0.133  0.89443    
year2014     0.007384   0.085918   0.086  0.93151    
year2015     0.164691   0.083937   1.962  0.04975 *  
year2016     0.226567   0.080112   2.828  0.00468 ** 
year2017     0.211735   0.079291   2.670  0.00758 ** 
year2018     0.181078   0.079704   2.272  0.02309 *  
year2019     0.218799   0.089074   2.456  0.01403 *  
rise        -2.691779   0.385087  -6.990 2.75e-12 ***
breakpt     -0.017529   0.011232  -1.561  0.11860    
tot.brk     -0.984249   0.034800 -28.283  < 2e-16 ***
h.brk2      -0.389623   0.041932  -9.292  < 2e-16 ***
loc          0.083007   0.011360   7.307 2.74e-13 ***
MPH         -0.095149   0.005519 -17.239  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 39587  on 277609  degrees of freedom
Residual deviance: 38191  on 277592  degrees of freedom
AIC: 38227

Number of Fisher Scoring iterations: 7

[1] "Predicted HRs:  3733 Actual HRs: 3734"

       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann          Effron 
          0.938           0.938           0.875           0.982           0.388           0.962           0.004 
McKelveyZavoina            Tjur             AIC             BIC          logLik         logLik0              G2 
          0.108           0.005       38226.797       38416.408      -19095.398     -308027.089      577863.382 

[1] "***** OUTPUT FOR FF *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 + 
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc), 
    ])

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.2249  -0.2004  -0.1809  -0.1632   3.2726  

Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  2.472246   0.211830  11.671  < 2e-16 ***
year2009     0.095921   0.033385   2.873  0.00406 ** 
year2010    -0.024734   0.034572  -0.715  0.47434    
year2011     0.013260   0.034607   0.383  0.70160    
year2012     0.126591   0.033388   3.791  0.00015 ***
year2013     0.031387   0.033968   0.924  0.35548    
year2014    -0.047886   0.035080  -1.365  0.17224    
year2015     0.138458   0.033218   4.168 3.07e-05 ***
year2016     0.239549   0.032395   7.395 1.42e-13 ***
year2017     0.397599   0.032109  12.383  < 2e-16 ***
year2018     0.313650   0.032256   9.724  < 2e-16 ***
year2019     0.481475   0.034619  13.908  < 2e-16 ***
rise         1.628400   0.400095   4.070 4.70e-05 ***
breakpt     -0.082688   0.013354  -6.192 5.94e-10 ***
tot.brk      0.166788   0.009440  17.668  < 2e-16 ***
h.brk2       0.194767   0.021264   9.159  < 2e-16 ***
loc          0.087070   0.004298  20.257  < 2e-16 ***
MPH         -0.080674   0.002249 -35.872  < 2e-16 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 232098  on 1334924  degrees of freedom
Residual deviance: 229702  on 1334907  degrees of freedom
AIC: 229738

Number of Fisher Scoring iterations: 7

[1] "Predicted HRs:  22961 Actual HRs: 22972"

       McFadden      McFaddenAdj         CoxSnell       Nagelkerke    AldrichNelson VeallZimmermann           Effron
          0.627            0.627            0.251            0.680            0.081           0.657            0.001
McKelveyZavoina             Tjur              AIC              BIC           logLik         logLik0               G2
          0.028            0.002       229738.372       229956.251      -114851.186     -308027.089       386351.807

[1] "***** OUTPUT FOR FT *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 + 
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc), 
    ])

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.6911  -0.1895  -0.1713  -0.1547   3.4701  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)  3.065613   0.419347   7.310 2.66e-13 ***
year2009     0.127537   0.061885   2.061  0.03932 *  
year2010     0.030807   0.062282   0.495  0.62086    
year2011    -0.015674   0.063517  -0.247  0.80508    
year2012     0.047887   0.061787   0.775  0.43832    
year2013     0.011218   0.061244   0.183  0.85467    
year2014    -0.176447   0.064118  -2.752  0.00592 ** 
year2015     0.061854   0.061215   1.010  0.31228    
year2016     0.238608   0.060966   3.914 9.09e-05 ***
year2017     0.330236   0.058672   5.628 1.82e-08 ***
year2018     0.204303   0.061180   3.339  0.00084 ***
year2019     0.324918   0.072613   4.475 7.65e-06 ***
rise         1.832715   1.759144   1.042  0.29749    
breakpt     -0.145950   0.029431  -4.959 7.08e-07 ***
tot.brk     -0.151517   0.017743  -8.540  < 2e-16 ***
h.brk2      -0.337188   0.043916  -7.678 1.61e-14 ***
loc          0.135347   0.009301  14.552  < 2e-16 ***
MPH         -0.071575   0.004468 -16.020  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 71051  on 446654  degrees of freedom
Residual deviance: 70334  on 446637  degrees of freedom
AIC: 70370

Number of Fisher Scoring iterations: 7

[1] "Predicted HRs:  6877 Actual HRs: 6880"
       McFadden      McFaddenAdj         CoxSnell       Nagelkerke    AldrichNelson VeallZimmermann           Effron
          0.886            0.886            0.705            0.943            0.271           0.917            0.001
McKelveyZavoina             Tjur              AIC              BIC           logLik         logLik0               G2
          0.029            0.002        70370.070        70568.242       -35167.035     -308027.089       545720.109

[1] "***** OUTPUT FOR SI *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 + 
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc), 
```

```
    ])

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.0999  -0.1887  -0.1700  -0.1529   3.3740  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept)  3.170700   0.444163   7.139 9.43e-13 ***
year2009     0.086668   0.055862   1.551  0.12079    
year2010    -0.005753   0.056290  -0.102  0.91859    
year2011     0.047291   0.057627   0.821  0.41185    
year2012     0.057140   0.060123   0.950  0.34191    
year2013    -0.058677   0.065752  -0.892  0.37218    
year2014    -0.179236   0.066987  -2.676  0.00746 ** 
year2015     0.036315   0.064705   0.561  0.57463    
year2016     0.153639   0.060829   2.526  0.01154 *  
year2017     0.290512   0.063972   4.541 5.59e-06 ***
year2018     0.117244   0.063412   1.849  0.06447 .  
year2019     0.432512   0.069752   6.201 5.62e-10 ***
rise        -0.471199   0.352706  -1.336  0.18156    
breakpt     -0.048457   0.010065  -4.814 1.48e-06 ***
tot.brk     -0.189543   0.020097  -9.431  < 2e-16 ***
h.brk2      -0.429238   0.048761  -8.803  < 2e-16 ***
loc          0.106973   0.008623  12.406  < 2e-16 ***
MPH         -0.070333   0.004753 -14.799  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 58885  on 374625  degrees of freedom
Residual deviance: 58279  on 374608  degrees of freedom
AIC: 58315

Number of Fisher Scoring iterations: 7

[1] "Predicted HRs:  5683 Actual HRs: 5685"
      McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann          Effron
         0.905           0.905           0.774           0.960           0.312           0.935           0.001
McKelveyZavoina            Tjur             AIC             BIC          logLik         logLik0              G2
         0.030           0.002       58314.634       58509.640      -29139.317     -308027.089      557775.545

[1] "***** OUTPUT FOR SL *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 + 
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc), 
    ])

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.3194  -0.1906  -0.1573  -0.1276   3.4707  

Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  5.1488342  0.3376560  15.249  < 2e-16 ***
year2009     0.0516044  0.0547351   0.943   0.3458    
year2010    -0.0180463  0.0556702  -0.324   0.7458    
year2011    -0.0007088  0.0546119  -0.013   0.9896    
year2012     0.0897572  0.0538852   1.666   0.0958 .  
year2013     0.0435433  0.0542943   0.802   0.4226    
year2014    -0.0240314  0.0568126  -0.423   0.6723    
year2015     0.0771808  0.0557280   1.385   0.1661    
year2016     0.2886037  0.0524866   5.499 3.83e-08 ***
year2017     0.2854631  0.0508892   5.610 2.03e-08 ***
year2018     0.0958461  0.0520617   1.841   0.0656 .  
year2019     0.3523881  0.0544249   6.475 9.50e-11 ***
rise         0.4576624  0.3781738   1.210   0.2262    
```

```
breakpt      -0.1169774  0.0082475 -14.183  < 2e-16 ***
tot.brk      -0.7778465  0.0178233 -43.642  < 2e-16 ***
h.brk2       -0.3202224  0.0336349  -9.521  < 2e-16 ***
loc           0.1134379  0.0071311  15.908  < 2e-16 ***
MPH          -0.0771562  0.0037408 -20.626  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 92371  on 623585  degrees of freedom
Residual deviance: 89888  on 623568  degrees of freedom
AIC: 89924

Number of Fisher Scoring iterations: 7

[1] "Predicted HRs:  8789 Actual HRs: 8790"
       McFadden      McFaddenAdj         CoxSnell       Nagelkerke    AldrichNelson VeallZimmermann           Effron
          0.854            0.854            0.570            0.908            0.204           0.884            0.003
McKelveyZavoina             Tjur              AIC              BIC           logLik         logLik0               G2
          0.084            0.004        89923.537        90127.715       -44943.768     -308027.089       526166.642
```

Appendix C: Generalized Linear Models, by pitch type, without year as factor

This set of models is identical to the first group in Appendix B, except with *year* removed as a factor from the model. The purpose of this is to examine the effect of pseudo-R^2 for pitching components only. See body of paper.

```
[1] "***** OUTPUT FOR CH *****"

Call:
glm(formula = HR ~ rise + breakpt + tot.brk + h.brk2 + loc +
    MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1241  -0.1436  -0.1161  -0.0867   5.5417

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.976939   0.319293   6.192 5.96e-10 ***
rise         3.881815   0.280877  13.820  < 2e-16 ***
breakpt     -0.260197   0.011017 -23.617  < 2e-16 ***
tot.brk     -0.724710   0.017603 -41.169  < 2e-16 ***
h.brk2      -0.393327   0.033538 -11.728  < 2e-16 ***
loc         -0.130551   0.008014 -16.291  < 2e-16 ***
MPH         -0.039553   0.003660 -10.807  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 77267  on 855627  degrees of freedom
Residual deviance: 74358  on 855621  degrees of freedom
AIC: 74372

Number of Fisher Scoring iterations: 8

[1] "Predicted HRs:  6590 Actual HRs: 6592"
    McFadden   McFaddenAdj      CoxSnell    Nagelkerke  AldrichNelson VeallZimmermann          Effron McKelveyZavoina            Tjur
       0.895         0.895         0.524         0.930          0.184           0.914           0.002           0.171           0.003
         AIC           BIC        logLik       logLik0             G2
   74372.084     74453.701    -37179.042   -354451.146     634544.208
[1] "***** OUTPUT FOR CU *****"

Call:
glm(formula = HR ~ rise + breakpt + tot.brk + h.brk2 + loc +
    MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3717  -0.1192  -0.0982  -0.0749   3.9219

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.791597   0.536550  12.658  < 2e-16 ***
rise        -4.367413   0.405060 -10.782  < 2e-16 ***
breakpt     -0.029632   0.011656  -2.542    0.011 *
tot.brk     -0.899982   0.035008 -25.708  < 2e-16 ***
h.brk2      -0.265318   0.041426  -6.405 1.51e-10 ***
loc         -0.103869   0.010202 -10.181  < 2e-16 ***
MPH         -0.096754   0.005489 -17.626  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 46499  on 697807  degrees of freedom
Residual deviance: 44898  on 697801  degrees of freedom
AIC: 44912
```

```
Number of Fisher Scoring iterations: 9

[1] "Predicted HRs:  3733 Actual HRs: 3734"
       McFadden      McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann          Effron McKelveyZavoina            Tjur
          0.937            0.937           0.614           0.962           0.224           0.951           0.002           0.193           0.002
            AIC              BIC          logLik         logLik0              G2
      44911.773        44991.963      -22448.887     -354451.146      664004.519
[1] "***** OUTPUT FOR FF *****"

Call:
glm(formula = HR ~ rise + breakpt + tot.brk + h.brk2 + loc +
    MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8546  -0.1352  -0.1260  -0.1159   4.0809

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.050011   0.211491   4.965 6.88e-07 ***
rise         3.850171   0.328798  11.710  < 2e-16 ***
breakpt     -0.186492   0.013150 -14.182  < 2e-16 ***
tot.brk     -0.063202   0.007980  -7.920 2.37e-15 ***
h.brk2       0.158773   0.021227   7.480 7.44e-14 ***
loc         -0.104740   0.004428 -23.656  < 2e-16 ***
MPH         -0.059707   0.002232 -26.752  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 267957  on 2900994  degrees of freedom
Residual deviance: 266270  on 2900988  degrees of freedom
AIC: 266284

Number of Fisher Scoring iterations: 9

[1] "Predicted HRs:  22961 Actual HRs: 22972"
       McFadden      McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann          Effron McKelveyZavoina            Tjur
          0.624            0.624           0.142           0.653           0.044           0.641           0.000           0.038           0.000
            AIC              BIC          logLik         logLik0              G2
     266284.043       266374.207     -133135.021     -354451.146      442632.250

[1] "***** OUTPUT FOR FT *****"

Call:
glm(formula = HR ~ rise + breakpt + tot.brk + h.brk2 + loc +
    MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4828  -0.1284  -0.1165  -0.1050   4.2855

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.481519   0.408591   3.626 0.000288 ***
rise         5.868617   1.158675   5.065 4.09e-07 ***
breakpt     -0.278361   0.026845 -10.369  < 2e-16 ***
tot.brk     -0.274276   0.015390 -17.822  < 2e-16 ***
h.brk2      -0.315531   0.043553  -7.245 4.33e-13 ***
loc         -0.094057   0.008062 -11.667  < 2e-16 ***
MPH         -0.053134   0.004318 -12.306  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 82123  on 994774  degrees of freedom
```

```
Residual deviance: 81371  on 994768  degrees of freedom
AIC: 81385

Number of Fisher Scoring iterations: 9

[1] "Predicted HRs:  6877 Actual HRs: 6880"
      McFadden    McFaddenAdj       CoxSnell     Nagelkerke   AldrichNelson VeallZimmermann          Effron McKelveyZavoina            Tjur
         0.885          0.885          0.468          0.918           0.161           0.904           0.001           0.049           0.001
           AIC            BIC         logLik        logLik0              G2
     81385.071      81467.743     -40685.536    -354451.146      627531.221

[1] "***** OUTPUT FOR SI *****"

Call:
glm(formula = HR ~ rise + breakpt + tot.brk + h.brk2 + loc +
    MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.2284  -0.1275  -0.1144  -0.1017   3.5640

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.482688   0.437197   3.391 0.000695 ***
rise         0.483931   0.304497   1.589 0.111997
breakpt     -0.083350   0.009320  -8.943  < 2e-16 ***
tot.brk     -0.296266   0.017750 -16.691  < 2e-16 ***
h.brk2      -0.406027   0.048287  -8.409  < 2e-16 ***
loc         -0.100257   0.008910 -11.252  < 2e-16 ***
MPH         -0.052395   0.004631 -11.313  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68139  on 842019  degrees of freedom
Residual deviance: 67487  on 842013  degrees of freedom
AIC: 67501

Number of Fisher Scoring iterations: 8

[1] "Predicted HRs:  5683 Actual HRs: 5685"
      McFadden    McFaddenAdj       CoxSnell     Nagelkerke   AldrichNelson VeallZimmermann          Effron McKelveyZavoina            Tjur
         0.905          0.905          0.533          0.937           0.188           0.923           0.001           0.039           0.001
           AIC            BIC         logLik        logLik0              G2
     67500.962      67582.466     -33743.481    -354451.146      641415.331

[1] "***** OUTPUT FOR SL *****"

Call:
glm(formula = HR ~ rise + breakpt + tot.brk + h.brk2 + loc +
    MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0062  -0.1346  -0.1109  -0.0871   3.6802

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.065497   0.330345  12.307  < 2e-16 ***
rise         1.616333   0.274394   5.891 3.85e-09 ***
breakpt     -0.176169   0.006782 -25.976  < 2e-16 ***
tot.brk     -0.743177   0.016919 -43.925  < 2e-16 ***
h.brk2      -0.270974   0.033485  -8.092 5.85e-16 ***
loc         -0.063309   0.006537  -9.685  < 2e-16 ***
MPH         -0.068377   0.003629 -18.843  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 105215  on 1290146  degrees of freedom
Residual deviance: 102181  on 1290140  degrees of freedom
AIC: 102195

Number of Fisher Scoring iterations: 8

[1] "Predicted HRs:  8789 Actual HRs: 8790"
      McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann          Effron McKelveyZavoina            Tjur
         0.856           0.856           0.375           0.887           0.125           0.874           0.001           0.132           0.002
           AIC             BIC          logLik         logLik0              G2
    102194.799      102279.290      -51090.399     -354451.146      606721.494
```

Appendix D: Generalized Linear Models, by pitch type, with year as numeric

In this appendix we show a general linear logistic regression model, of the form shown in the function call. The variable "year" was entered as a numeric variable. Six models are shown, one for each pitch type. The striking feature is the small p-values of the pitch components, as well as the consistent 2016, 2017, and 2019 years as varying from the rest. The signs of the coefficients are as expected.

```
> for (i in 1:6) run.GLM(All.pit, pitch.type=pitch.types[i])

[1] "***** OUTPUT FOR CH *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1769  -0.1437  -0.1158  -0.0865   5.5153

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -41.271490   7.566301  -5.455 4.91e-08 ***
year          0.021587   0.003784   5.704 1.17e-08 ***
rise          3.893461   0.282020  13.806  < 2e-16 ***
breakpt      -0.260722   0.011015 -23.669  < 2e-16 ***
tot.brk      -0.722060   0.017616 -40.989  < 2e-16 ***
h.brk2       -0.413794   0.033794 -12.244  < 2e-16 ***
loc          -0.130187   0.008018 -16.238  < 2e-16 ***
MPH          -0.042311   0.003737 -11.323  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 77267  on 855627  degrees of freedom
Residual deviance: 74334  on 855620  degrees of freedom
AIC: 74350

Number of Fisher Scoring iterations: 8

[1] "Predicted HRs:  6590 Actual HRs: 6592"
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson
VeallZimmermann          Effron
          0.895           0.895           0.524           0.930           0.184
0.914          0.002
McKelveyZavoina            Tjur             AIC             BIC          logLik
logLik0              G2
          0.172           0.003       74349.675       74442.952      -37166.838     -
354451.146       634568.617
[1] "***** OUTPUT FOR CU *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3902  -0.1192  -0.0979  -0.0746   3.9077

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -51.472507   9.927262  -5.185 2.16e-07 ***
year          0.029027   0.004945   5.869 4.37e-09 ***
rise         -4.349784   0.405511 -10.727  < 2e-16 ***
breakpt      -0.029969   0.011684  -2.565   0.0103 *
tot.brk      -0.890858   0.035166 -25.333  < 2e-16 ***
h.brk2       -0.259759   0.041497  -6.260 3.86e-10 ***
loc          -0.104788   0.010206 -10.268  < 2e-16 ***
MPH          -0.100261   0.005571 -17.998  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 46499  on 697807  degrees of freedom
Residual deviance: 44873  on 697800  degrees of freedom
AIC: 44889

Number of Fisher Scoring iterations: 9

[1] "Predicted HRs:  3733 Actual HRs: 3734"
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson
VeallZimmermann          Effron
          0.937           0.937           0.614           0.962           0.224
0.951          0.002
McKelveyZavoina            Tjur             AIC             BIC          logLik
logLik0              G2
          0.193           0.002       44889.360       44981.005      -22436.680     -
354451.146      664028.933
[1] "***** OUTPUT FOR FF *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7505  -0.1360  -0.1253  -0.1142   4.0181

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -71.709075   4.041011 -17.745  < 2e-16 ***
year          0.036335   0.002021  17.977  < 2e-16 ***
rise          3.738136   0.329954  11.329  < 2e-16 ***
breakpt      -0.183562   0.013118 -13.993  < 2e-16 ***
tot.brk      -0.050867   0.007988  -6.368 1.92e-10 ***
h.brk2        0.142530   0.021212   6.719 1.82e-11 ***
loc          -0.103282   0.004432 -23.306  < 2e-16 ***
MPH          -0.064934   0.002277 -28.512  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

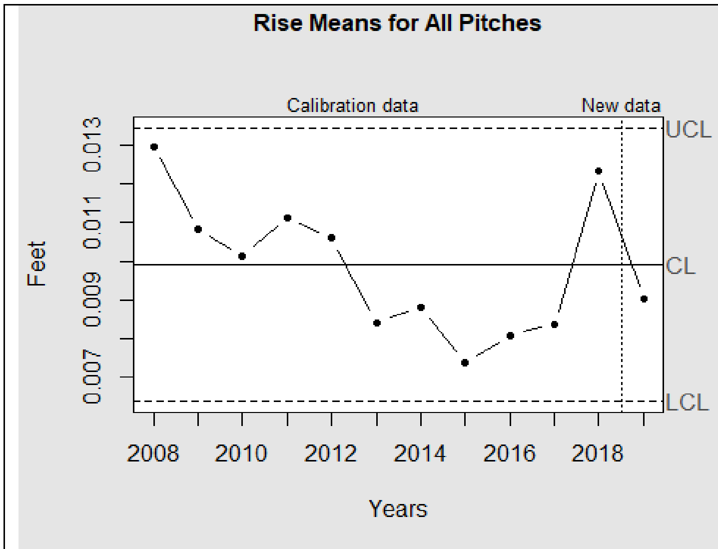
```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 267957  on 2900994  degrees of freedom
Residual deviance: 266000  on 2900987  degrees of freedom
AIC: 266016

Number of Fisher Scoring iterations: 9

[1] "Predicted HRs:  22961 Actual HRs: 22972"
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson
VeallZimmermann          Effron
          0.625           0.625           0.142           0.653           0.044
0.641          0.001
McKelveyZavoina            Tjur             AIC             BIC          logLik
logLik0             G2
          0.041           0.001      266016.337      266119.382     -133000.169      -
354451.146      442901.955
[1] "***** OUTPUT FOR FT *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4886  -0.1285  -0.1163  -0.1044   4.2595

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -49.558995   7.637616  -6.489 8.65e-11 ***
year          0.025464   0.003823   6.660 2.74e-11 ***
rise          5.905748   1.160731   5.088 3.62e-07 ***
breakpt      -0.277291   0.026667 -10.398  < 2e-16 ***
tot.brk      -0.269153   0.015399 -17.479  < 2e-16 ***
h.brk2       -0.318370   0.043595  -7.303 2.82e-13 ***
loc          -0.093172   0.008066 -11.551  < 2e-16 ***
MPH          -0.056211   0.004411 -12.744  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 82123  on 994774  degrees of freedom
Residual deviance: 81341  on 994767  degrees of freedom
AIC: 81357

Number of Fisher Scoring iterations: 9

[1] "Predicted HRs:  6877 Actual HRs: 6880"
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson
VeallZimmermann          Effron
          0.885           0.885           0.468           0.918           0.161
0.904          0.001
McKelveyZavoina            Tjur             AIC             BIC          logLik
logLik0             G2
```

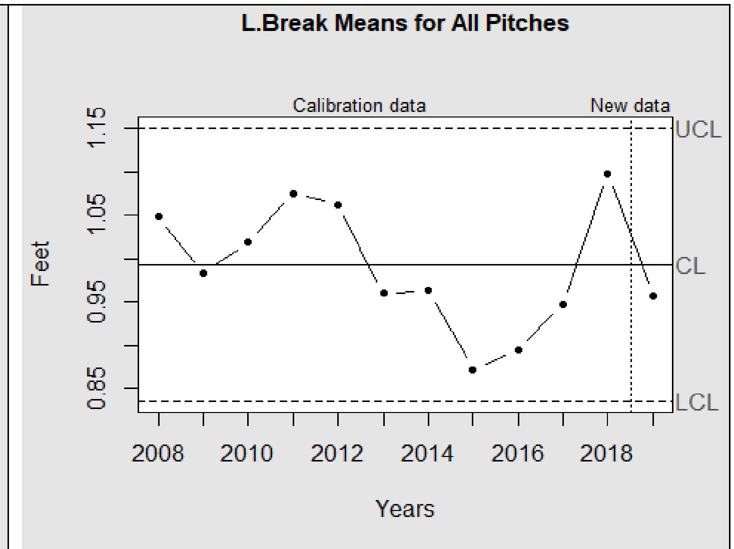
```
          0.050           0.001       81356.938       81451.420      -40670.469      -
354451.146      627561.354
[1] "***** OUTPUT FOR SI *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.2396  -0.1275  -0.1143  -0.1014   3.5894

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -35.343254   7.970947  -4.434 9.25e-06 ***
year          0.018354   0.003983   4.609 4.05e-06 ***
rise          0.507536   0.305603   1.661   0.0968 .
breakpt      -0.083909   0.009347  -8.977  < 2e-16 ***
tot.brk      -0.289976   0.017773 -16.316  < 2e-16 ***
h.brk2       -0.403924   0.048357  -8.353  < 2e-16 ***
loc          -0.099466   0.008916 -11.155  < 2e-16 ***
MPH          -0.054326   0.004730 -11.484  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68139  on 842019  degrees of freedom
Residual deviance: 67474  on 842012  degrees of freedom
AIC: 67490

Number of Fisher Scoring iterations: 8

[1] "Predicted HRs:  5683 Actual HRs: 5685"
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson
VeallZimmermann          Effron
          0.905           0.905           0.533           0.937           0.188
0.923          0.001
McKelveyZavoina            Tjur             AIC             BIC          logLik
logLik0              G2
          0.040           0.001       67489.692       67582.840      -33736.846      -
354451.146      641428.601
[1] "***** OUTPUT FOR SL *****"

Call:
glm(formula = HR ~ year + rise + breakpt + tot.brk + h.brk2 +
    loc + MPH, family = binomial(), data = All.pit[is.finite(All.pit$loc),
    ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0050  -0.1346  -0.1105  -0.0865   3.7097

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -56.589523   6.324563  -8.948  < 2e-16 ***
year          0.030241   0.003158   9.576  < 2e-16 ***
```

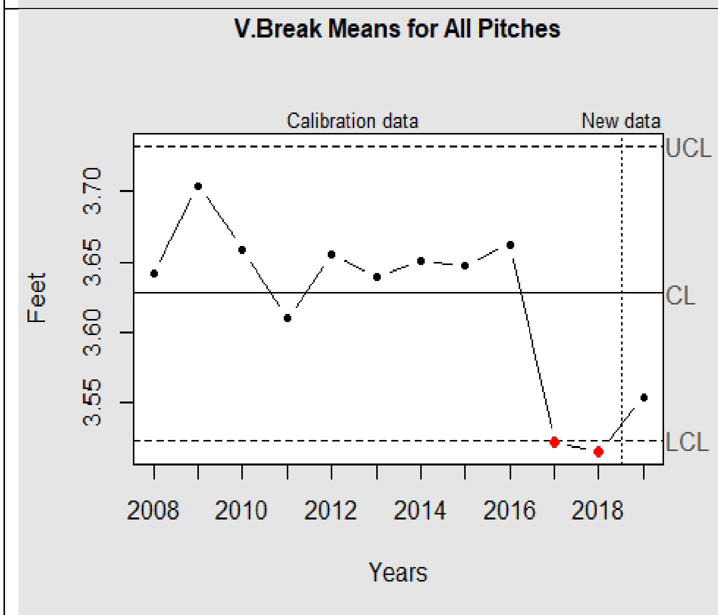
```
rise          1.596863   0.272766   5.854 4.79e-09 ***
breakpt      -0.176384   0.006785 -25.995  < 2e-16 ***
tot.brk      -0.741065   0.016952 -43.716  < 2e-16 ***
h.brk2       -0.278186   0.033488  -8.307  < 2e-16 ***
loc          -0.063775   0.006540  -9.752  < 2e-16 ***
MPH          -0.071705   0.003681 -19.480  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 105215  on 1290146  degrees of freedom
Residual deviance: 102109  on 1290139  degrees of freedom
AIC: 102125

Number of Fisher Scoring iterations: 8

[1] "Predicted HRs:  8789 Actual HRs: 8790"
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson
VeallZimmermann          Effron
          0.856           0.856           0.375           0.888           0.125
0.874          0.002
McKelveyZavoina            Tjur             AIC             BIC          logLik
logLik0             G2
          0.134           0.002      102124.945      102221.508      -51054.473     -
354451.146      606793.347
```

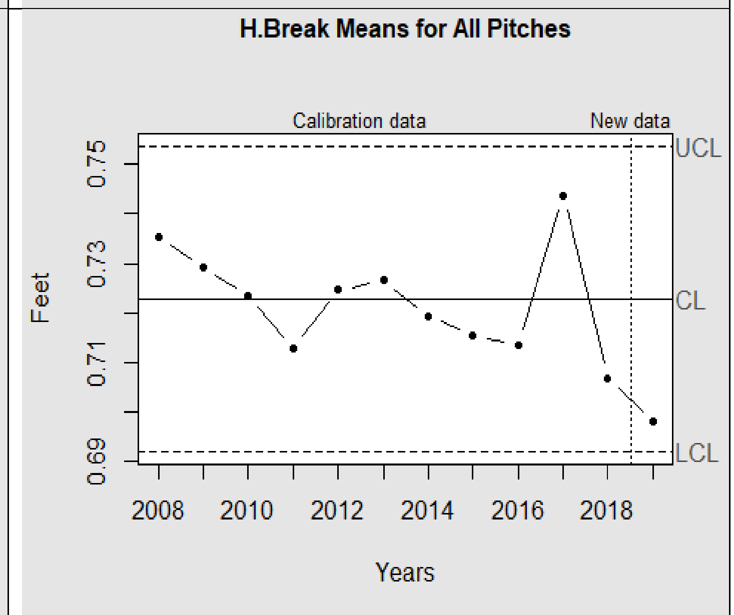
Appendix E: Control Charts showing changes in components for pitches swung on

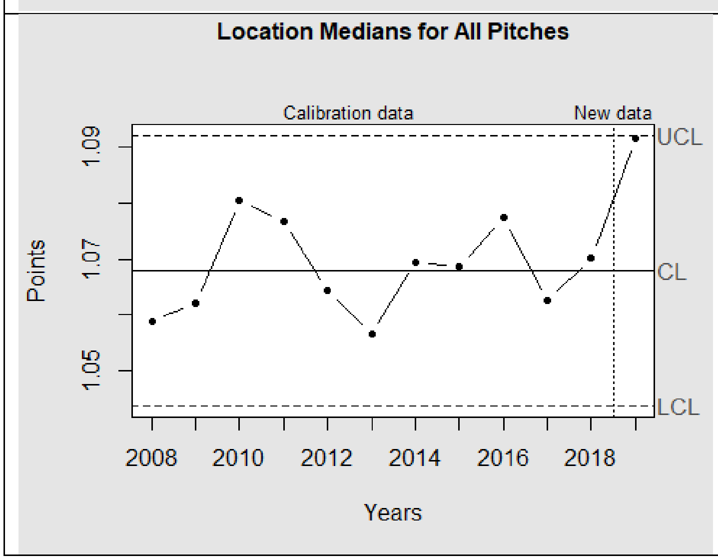
These control charts are the same as in the body of the paper, except they are calculated on pitches swung on only.

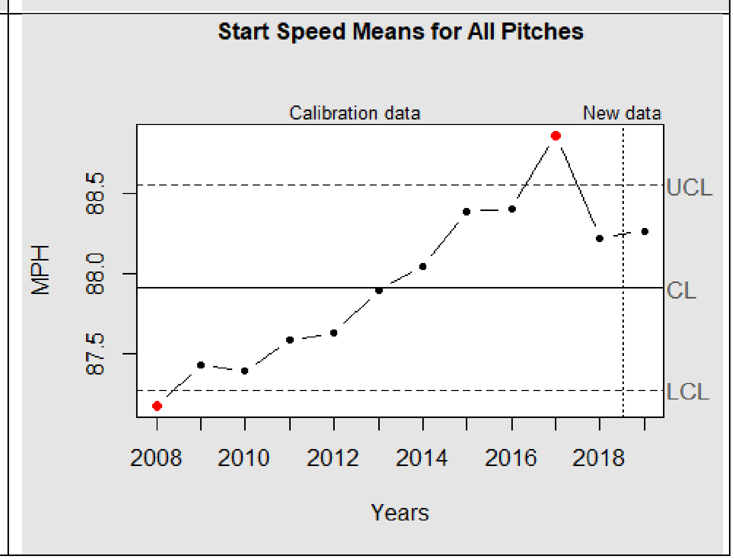












Appendix F: Comparison of our zone numbers with PITCHf/x and Statcast

Below is a comparison of the counts, by zone, of the 2018 pitches between PITCHf/x, Statcast, and our zone algorithms. The data for all years were obtained using R's MLBGameday package as of July 18, 2019. Our R function for obtaining the *Wilson* zones is below.

	1	2	3	4	5	6	7	8	9	11	12	13	14
PITCHf/x	24816	28014	23456	35228	40758	36760	30289	35972	34166	92528	74000	115461	156494
Wilson	27816	33203	25844	39584	48251	41125	33826	42545	38890	83408	65733	105613	144858
Statcast	29041	34138	26541	44243	53403	45918	35990	45072	41767	77608	60658	97776	135785

```
make.zone2 <- function(data)
#Construct strike zone of 1,2,3,...,9,11,12,13,14 to mimic MLB zones.
#Differs from make.zone() in that I honed the 11,12,13,14 zones more accurately.
#Occasion was PITCHf/x and StatCast zones differ + 2019 PITCHf/x didn't come
#Data needs to have: px,pz,sz_top,sz_bot
#Reference for strike zone:
#https://baseballwithr.wordpress.com/2015/02/17/conceptualizing-the-mlb-strike-zone-using-pitchfx-data/
#https://tht.fangraphs.com/the-2017-strike-zone/
#Edge of strikze zone: ((1.57*2 + 17) / 12) / 2 = 0.8391667
#Horizontal is x0,x1,x2,x3, left to right and Vertical is z0,z1,z2,z3, bottom to top
{

  edge = 0.8391667
  x2 = 2*edge/3 - (edge/3) #right vertical line
  x1 = -x2                 #left vertical line
  x0 = -edge               #left vertical edge of strike zone
  x3 = edge                #right vertical edge of strike zone
  z0 = data$sz_bot
  z3 = data$sz_top
  z1 = (z3-z0)/3 + z0
  z2 = (z3-z0)/3 + z1
  z1.5 = (z0+z3)/2

  #Safe calculation of the zone.  For fast and interesting approach, see
plot.zone()
  px = data$px; pz=data$pz; sz_top=data$sz_top; sz_bot = data$sz_bot
  zone2 = ifelse(px>x0 & px<=x1 & pz>z0 & pz<=z1, 7, NA)
  zone2 = ifelse(px>x1 & px<=x2 & pz>z0 & pz<=z1, 8, zone2)
  zone2 = ifelse(px>x2 & px<=x3 & pz>z0 & pz<=z1, 9, zone2)
  zone2 = ifelse(px>x0 & px<=x1 & pz>z1 & pz<=z2, 4, zone2)
  zone2 = ifelse(px>x1 & px<=x2 & pz>z1 & pz<=z2, 5, zone2)
  zone2 = ifelse(px>x2 & px<=x3 & pz>z1 & pz<=z2, 6, zone2)
  zone2 = ifelse(px>x0 & px<=x1 & pz>z2 & pz<=z3, 1, zone2)
  zone2 = ifelse(px>x1 & px<=x2 & pz>z2 & pz<=z3, 2, zone2)
  zone2 = ifelse(px>x2 & px<=x3 & pz>z2 & pz<=z3, 3, zone2)
  zone2 = ifelse((px<x0 & pz>=z1.5) | (px<=0 & pz>=z3), 11, zone2)
  zone2 = ifelse((px>=x3 & pz>=z1.5) | (px>=0 & pz>=z3), 12, zone2)
```

```
  zone2 = ifelse((px<x0 & pz<z1.5) | (px<0 & pz<z0), 13, zone2)
  zone2 = ifelse((px>=x3 & pz<z1.5) | (px>=0 & pz<z0), 14, zone2)

  return(zone2)
}
```

Appendix G: Pitcher-Batter Handedness Splits

This appendix documents the same all-zone graphs and tables shown in the body, except for handedness splits. The control charts are given in Appendix H.

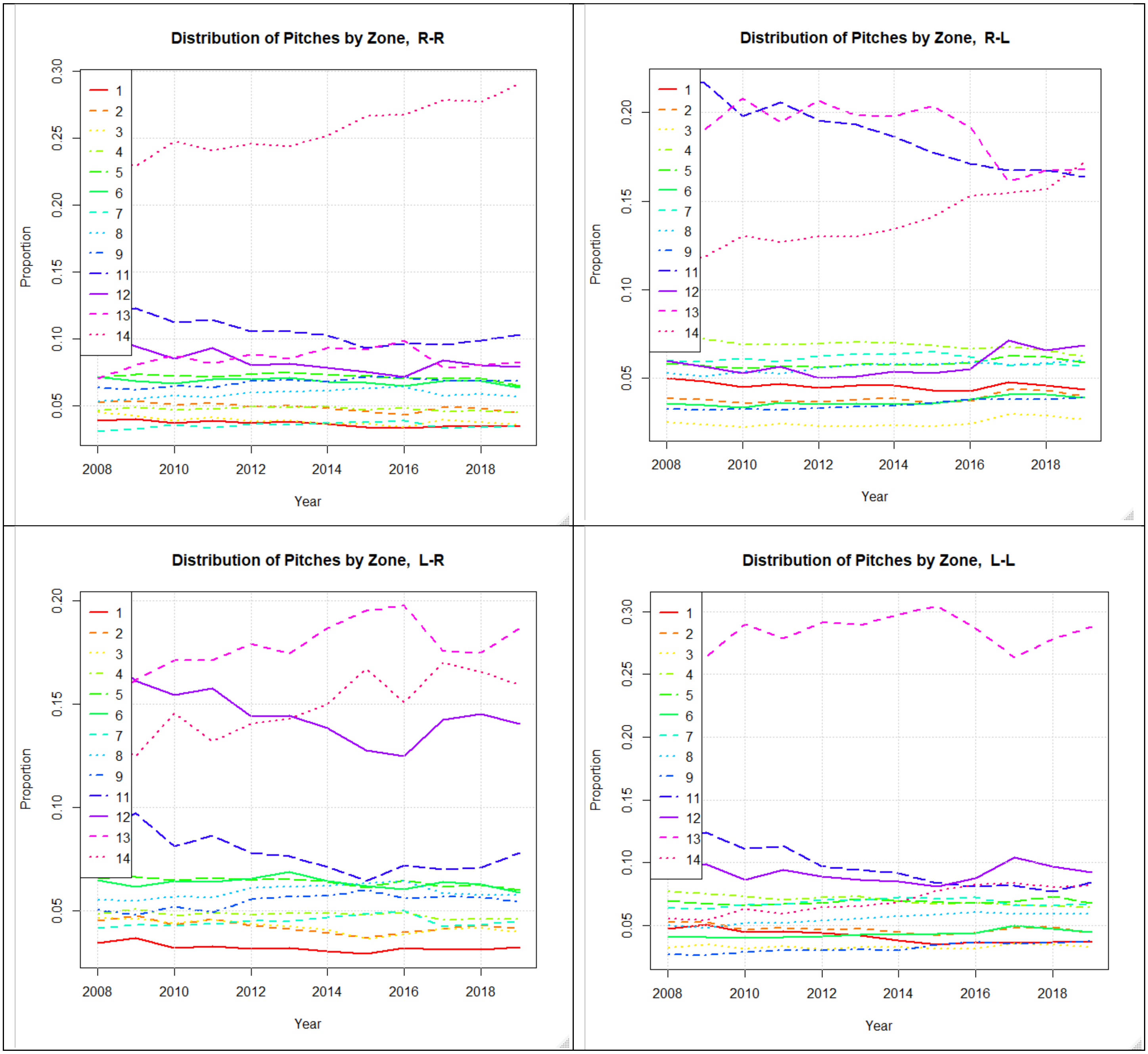


Figure G.1. Distribution of Pitches by Zone, Split by Pitcher-Batter Handedness

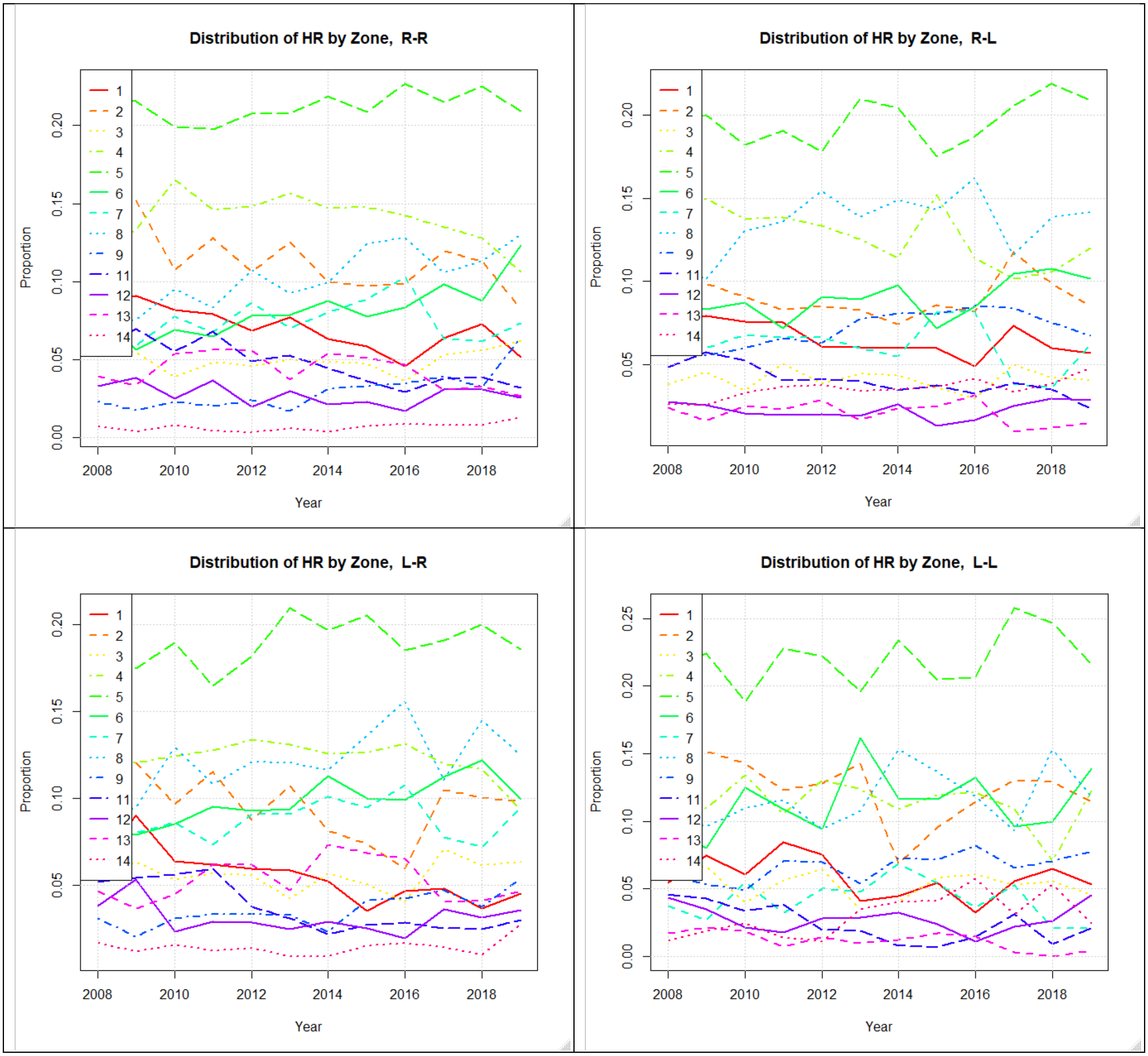


Figure G.2. Distribution of Home Runs by Zone, Split by Pitcher-Batter Handedness

Figure G.3. Proportion of Home Runs in Each Zone, Split by Pitcher-Batter Handedness

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.038	0.052	0.045	0.046	0.070	0.072	0.031	0.053	0.064	0.118	0.106	0.070	0.235
2009	0.040	0.053	0.043	0.048	0.073	0.068	0.032	0.055	0.061	0.123	0.094	0.080	0.229
2010	0.037	0.050	0.038	0.047	0.072	0.066	0.035	0.057	0.065	0.112	0.085	0.088	0.247
2011	0.038	0.052	0.041	0.048	0.071	0.069	0.033	0.056	0.063	0.114	0.093	0.081	0.241
2012	0.037	0.049	0.038	0.049	0.073	0.069	0.036	0.060	0.068	0.106	0.080	0.089	0.246
2013	0.038	0.050	0.039	0.049	0.075	0.071	0.035	0.060	0.069	0.105	0.081	0.085	0.244
2014	0.036	0.048	0.036	0.049	0.073	0.068	0.037	0.061	0.068	0.102	0.078	0.093	0.251
2015	0.034	0.046	0.036	0.047	0.072	0.067	0.037	0.063	0.072	0.093	0.075	0.092	0.267
2016	0.033	0.043	0.034	0.048	0.070	0.065	0.039	0.064	0.071	0.096	0.071	0.099	0.267
2017	0.034	0.049	0.039	0.046	0.070	0.068	0.033	0.057	0.068	0.095	0.083	0.078	0.278
2018	0.035	0.048	0.037	0.046	0.070	0.068	0.034	0.058	0.068	0.098	0.080	0.080	0.277
2019	0.034	0.044	0.035	0.045	0.065	0.063	0.034	0.056	0.068	0.103	0.079	0.083	0.290

Table G.1. Distribution of Pitches, R-R

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.086	0.141	0.056	0.118	0.221	0.082	0.064	0.072	0.023	0.058	0.033	0.039	0.007
2009	0.091	0.152	0.055	0.134	0.216	0.056	0.059	0.075	0.017	0.069	0.038	0.034	0.004
2010	0.082	0.108	0.039	0.165	0.199	0.069	0.078	0.095	0.023	0.055	0.025	0.054	0.008
2011	0.079	0.128	0.048	0.146	0.198	0.065	0.067	0.084	0.020	0.068	0.037	0.056	0.004
2012	0.069	0.106	0.046	0.148	0.208	0.078	0.086	0.108	0.024	0.049	0.019	0.056	0.003
2013	0.077	0.125	0.050	0.157	0.207	0.078	0.070	0.092	0.017	0.052	0.030	0.037	0.006
2014	0.063	0.100	0.048	0.147	0.219	0.088	0.080	0.099	0.031	0.044	0.021	0.054	0.004
2015	0.058	0.097	0.047	0.148	0.209	0.078	0.088	0.124	0.033	0.036	0.023	0.051	0.007
2016	0.046	0.099	0.036	0.142	0.227	0.083	0.103	0.128	0.035	0.029	0.017	0.047	0.008
2017	0.064	0.120	0.053	0.135	0.215	0.098	0.063	0.105	0.039	0.038	0.031	0.030	0.008
2018	0.073	0.113	0.056	0.128	0.225	0.088	0.061	0.113	0.032	0.039	0.031	0.033	0.008
2019	0.051	0.082	0.062	0.106	0.210	0.123	0.074	0.130	0.064	0.032	0.025	0.027	0.013

Table G.2. Distribution of Home Runs, R-R

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.015	0.019	0.009	0.018	0.022	0.008	0.014	0.009	0.002	0.003	0.002	0.004	0
2009	0.016	0.021	0.009	0.020	0.021	0.006	0.013	0.010	0.002	0.004	0.003	0.003	0
2010	0.015	0.015	0.007	0.024	0.019	0.007	0.015	0.011	0.002	0.003	0.002	0.004	0
2011	0.014	0.016	0.008	0.020	0.018	0.006	0.014	0.010	0.002	0.004	0.003	0.005	0
2012	0.013	0.015	0.008	0.021	0.020	0.008	0.017	0.013	0.002	0.003	0.002	0.004	0
2013	0.013	0.016	0.008	0.021	0.018	0.007	0.013	0.010	0.002	0.003	0.002	0.003	0
2014	0.011	0.013	0.008	0.018	0.019	0.008	0.013	0.010	0.003	0.003	0.002	0.004	0
2015	0.012	0.015	0.009	0.022	0.021	0.008	0.017	0.014	0.003	0.003	0.002	0.004	0
2016	0.011	0.018	0.009	0.024	0.026	0.010	0.021	0.016	0.004	0.002	0.002	0.004	0
2017	0.016	0.021	0.012	0.025	0.027	0.012	0.017	0.016	0.005	0.003	0.003	0.003	0
2018	0.016	0.019	0.012	0.022	0.025	0.010	0.015	0.015	0.004	0.003	0.003	0.003	0
2019	0.014	0.017	0.016	0.022	0.030	0.018	0.020	0.021	0.009	0.003	0.003	0.003	0

Table G.3. Proportion of Home Runs, R-R

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.050	0.039	0.025	0.073	0.058	0.036	0.060	0.053	0.033	0.216	0.060	0.182	0.116
2009	0.048	0.038	0.024	0.072	0.057	0.035	0.060	0.051	0.032	0.217	0.057	0.190	0.119
2010	0.045	0.036	0.023	0.069	0.056	0.034	0.061	0.054	0.033	0.198	0.053	0.208	0.131
2011	0.047	0.037	0.024	0.069	0.057	0.037	0.059	0.053	0.032	0.205	0.057	0.195	0.127
2012	0.045	0.037	0.023	0.070	0.056	0.036	0.063	0.056	0.033	0.196	0.050	0.207	0.130
2013	0.046	0.038	0.023	0.071	0.058	0.036	0.064	0.058	0.034	0.193	0.051	0.198	0.130
2014	0.046	0.039	0.024	0.070	0.058	0.036	0.064	0.058	0.035	0.186	0.054	0.198	0.134
2015	0.043	0.036	0.023	0.069	0.057	0.036	0.065	0.058	0.037	0.178	0.053	0.204	0.141
2016	0.043	0.037	0.024	0.067	0.059	0.038	0.062	0.059	0.038	0.171	0.055	0.192	0.153
2017	0.048	0.044	0.030	0.068	0.063	0.041	0.057	0.057	0.038	0.167	0.072	0.161	0.155
2018	0.046	0.043	0.029	0.066	0.062	0.041	0.058	0.058	0.038	0.167	0.066	0.168	0.157
2019	0.044	0.040	0.027	0.062	0.059	0.039	0.057	0.060	0.039	0.164	0.068	0.168	0.172

Table G.4. Distribution of Pitches, R-L

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.074	0.094	0.038	0.142	0.197	0.085	0.060	0.115	0.067	0.049	0.028	0.024	0.027
2009	0.079	0.099	0.046	0.150	0.200	0.083	0.060	0.101	0.056	0.058	0.026	0.016	0.026
2010	0.076	0.091	0.035	0.137	0.182	0.087	0.068	0.131	0.060	0.053	0.021	0.025	0.033
2011	0.075	0.083	0.051	0.139	0.191	0.072	0.066	0.136	0.066	0.041	0.020	0.023	0.037
2012	0.061	0.086	0.039	0.134	0.178	0.091	0.067	0.154	0.063	0.041	0.020	0.029	0.038
2013	0.061	0.083	0.045	0.125	0.210	0.089	0.060	0.139	0.077	0.040	0.019	0.017	0.034
2014	0.060	0.074	0.044	0.114	0.204	0.098	0.055	0.149	0.081	0.035	0.026	0.024	0.036
2015	0.060	0.086	0.036	0.152	0.175	0.072	0.082	0.143	0.080	0.038	0.013	0.025	0.037
2016	0.049	0.082	0.030	0.114	0.187	0.085	0.082	0.162	0.085	0.033	0.017	0.032	0.042
2017	0.073	0.118	0.050	0.102	0.206	0.105	0.039	0.116	0.084	0.039	0.025	0.010	0.034
2018	0.060	0.099	0.042	0.106	0.219	0.108	0.036	0.138	0.076	0.035	0.030	0.012	0.039
2019	0.057	0.085	0.041	0.120	0.209	0.101	0.062	0.141	0.067	0.024	0.029	0.015	0.048

Table G.5. Distribution of Home Runs, R-L

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.010	0.017	0.010	0.013	0.023	0.016	0.007	0.015	0.014	0.002	0.003	0.001	0.002
2009	0.011	0.017	0.013	0.014	0.024	0.016	0.007	0.013	0.012	0.002	0.003	0.001	0.001
2010	0.011	0.016	0.010	0.013	0.021	0.016	0.007	0.015	0.012	0.002	0.003	0.001	0.002
2011	0.010	0.014	0.013	0.013	0.021	0.012	0.007	0.016	0.013	0.001	0.002	0.001	0.002
2012	0.009	0.016	0.011	0.013	0.022	0.017	0.007	0.019	0.013	0.001	0.003	0.001	0.002
2013	0.009	0.015	0.013	0.012	0.024	0.017	0.006	0.016	0.015	0.001	0.003	0.001	0.002
2014	0.008	0.011	0.011	0.009	0.020	0.016	0.005	0.015	0.013	0.001	0.003	0.001	0.002
2015	0.010	0.017	0.011	0.016	0.022	0.014	0.009	0.018	0.016	0.002	0.002	0.001	0.002
2016	0.009	0.017	0.010	0.013	0.025	0.017	0.010	0.021	0.017	0.001	0.002	0.001	0.002
2017	0.013	0.023	0.015	0.013	0.029	0.022	0.006	0.018	0.019	0.002	0.003	0.001	0.002
2018	0.010	0.018	0.012	0.013	0.028	0.021	0.005	0.019	0.016	0.002	0.004	0.001	0.002
2019	0.012	0.019	0.014	0.018	0.032	0.024	0.010	0.021	0.016	0.001	0.004	0.001	0.003

Table G.6. Proportion of Home Runs, R-L

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.034	0.045	0.046	0.049	0.066	0.065	0.041	0.055	0.051	0.088	0.177	0.147	0.136
2009	0.037	0.048	0.046	0.051	0.066	0.062	0.043	0.055	0.048	0.097	0.161	0.162	0.125
2010	0.032	0.043	0.044	0.048	0.065	0.065	0.043	0.057	0.052	0.081	0.155	0.171	0.146
2011	0.033	0.046	0.046	0.049	0.066	0.064	0.044	0.056	0.049	0.086	0.158	0.171	0.132
2012	0.032	0.043	0.044	0.048	0.065	0.066	0.045	0.061	0.055	0.078	0.144	0.179	0.140
2013	0.032	0.041	0.042	0.049	0.066	0.069	0.045	0.062	0.057	0.077	0.145	0.174	0.143
2014	0.030	0.039	0.041	0.049	0.064	0.065	0.047	0.062	0.057	0.071	0.139	0.187	0.150
2015	0.029	0.037	0.037	0.048	0.061	0.062	0.048	0.063	0.060	0.064	0.128	0.195	0.167
2016	0.032	0.040	0.038	0.049	0.065	0.061	0.050	0.064	0.056	0.072	0.125	0.198	0.151
2017	0.031	0.041	0.041	0.046	0.062	0.064	0.042	0.058	0.057	0.070	0.142	0.176	0.170
2018	0.031	0.043	0.042	0.046	0.062	0.063	0.043	0.057	0.056	0.071	0.145	0.175	0.166
2019	0.032	0.041	0.040	0.046	0.060	0.059	0.045	0.058	0.054	0.078	0.140	0.187	0.159

Table G.7. Distribution of Pitches, L-R

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.064	0.111	0.056	0.131	0.197	0.082	0.086	0.089	0.031	0.052	0.038	0.047	0.017
2009	0.090	0.121	0.063	0.121	0.175	0.079	0.080	0.095	0.021	0.054	0.053	0.037	0.012
2010	0.064	0.097	0.053	0.124	0.189	0.085	0.086	0.130	0.031	0.056	0.023	0.045	0.016
2011	0.062	0.115	0.057	0.128	0.165	0.095	0.073	0.108	0.034	0.060	0.029	0.062	0.013
2012	0.060	0.088	0.056	0.134	0.182	0.093	0.091	0.121	0.033	0.038	0.029	0.062	0.014
2013	0.059	0.107	0.042	0.131	0.209	0.094	0.091	0.121	0.033	0.031	0.025	0.047	0.009
2014	0.052	0.081	0.057	0.126	0.197	0.113	0.101	0.116	0.023	0.022	0.029	0.073	0.009
2015	0.035	0.074	0.051	0.126	0.205	0.100	0.095	0.136	0.041	0.027	0.025	0.069	0.015
2016	0.047	0.059	0.041	0.132	0.185	0.099	0.108	0.156	0.042	0.029	0.020	0.065	0.017
2017	0.048	0.105	0.071	0.120	0.191	0.112	0.078	0.111	0.047	0.025	0.036	0.041	0.014
2018	0.037	0.100	0.062	0.117	0.200	0.122	0.072	0.145	0.038	0.025	0.032	0.041	0.010
2019	0.046	0.098	0.064	0.094	0.186	0.100	0.095	0.125	0.054	0.030	0.036	0.047	0.028

Table G.8. Distribution of Home Runs, L-R

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.013	0.017	0.009	0.019	0.021	0.009	0.015	0.011	0.004	0.004	0.002	0.002	0.001
2009	0.018	0.019	0.010	0.018	0.019	0.009	0.014	0.013	0.003	0.004	0.002	0.002	0.001
2010	0.013	0.014	0.008	0.017	0.019	0.008	0.013	0.015	0.004	0.004	0.001	0.002	0.001
2011	0.013	0.018	0.009	0.018	0.018	0.010	0.012	0.013	0.005	0.005	0.001	0.003	0.001
2012	0.014	0.016	0.010	0.021	0.022	0.011	0.015	0.015	0.005	0.004	0.002	0.003	0.001
2013	0.012	0.018	0.007	0.018	0.021	0.009	0.014	0.013	0.004	0.003	0.001	0.002	0.000
2014	0.011	0.014	0.009	0.017	0.020	0.011	0.014	0.012	0.003	0.002	0.001	0.003	0.000
2015	0.009	0.015	0.010	0.019	0.025	0.012	0.014	0.016	0.005	0.003	0.001	0.003	0.001
2016	0.013	0.013	0.009	0.023	0.025	0.014	0.019	0.021	0.007	0.003	0.001	0.003	0.001
2017	0.013	0.022	0.015	0.023	0.027	0.015	0.016	0.016	0.007	0.003	0.002	0.002	0.001
2018	0.009	0.018	0.011	0.020	0.025	0.015	0.013	0.019	0.005	0.003	0.002	0.002	0.000
2019	0.014	0.023	0.016	0.020	0.030	0.017	0.020	0.021	0.010	0.004	0.003	0.002	0.002

Table G.9. Proportion of Home Runs, L-R

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.047	0.053	0.032	0.077	0.070	0.042	0.064	0.050	0.027	0.119	0.096	0.268	0.056
2009	0.051	0.052	0.035	0.075	0.067	0.041	0.063	0.048	0.026	0.124	0.099	0.263	0.054
2010	0.045	0.047	0.032	0.073	0.066	0.040	0.066	0.052	0.029	0.111	0.087	0.289	0.063
2011	0.046	0.048	0.034	0.070	0.068	0.041	0.067	0.052	0.030	0.113	0.094	0.278	0.059
2012	0.044	0.047	0.031	0.072	0.068	0.041	0.071	0.054	0.031	0.097	0.089	0.291	0.065
2013	0.042	0.047	0.033	0.073	0.071	0.043	0.070	0.055	0.031	0.095	0.086	0.289	0.065
2014	0.038	0.045	0.033	0.069	0.070	0.043	0.072	0.058	0.030	0.091	0.085	0.297	0.069
2015	0.035	0.042	0.031	0.067	0.068	0.043	0.071	0.058	0.035	0.084	0.081	0.304	0.078
2016	0.037	0.044	0.032	0.068	0.068	0.044	0.073	0.061	0.037	0.081	0.088	0.286	0.082
2017	0.037	0.048	0.036	0.067	0.069	0.050	0.066	0.060	0.035	0.082	0.104	0.263	0.084
2018	0.037	0.049	0.035	0.065	0.073	0.047	0.066	0.059	0.036	0.077	0.097	0.278	0.080
2019	0.037	0.044	0.033	0.065	0.068	0.045	0.067	0.059	0.037	0.085	0.092	0.287	0.082

Table G.10. Distribution of Pitches, L-L

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.055	0.152	0.060	0.106	0.216	0.095	0.037	0.101	0.060	0.046	0.043	0.017	0.011
2009	0.075	0.152	0.067	0.109	0.224	0.080	0.027	0.096	0.053	0.043	0.035	0.021	0.019
2010	0.061	0.143	0.040	0.134	0.188	0.125	0.055	0.109	0.049	0.033	0.021	0.018	0.024
2011	0.084	0.123	0.056	0.105	0.228	0.109	0.032	0.116	0.070	0.039	0.018	0.007	0.014
2012	0.075	0.128	0.064	0.131	0.222	0.094	0.050	0.094	0.069	0.019	0.028	0.014	0.011
2013	0.041	0.142	0.035	0.123	0.196	0.161	0.047	0.108	0.054	0.019	0.028	0.009	0.035
2014	0.044	0.069	0.040	0.109	0.234	0.117	0.069	0.153	0.073	0.008	0.032	0.012	0.040
2015	0.055	0.096	0.058	0.119	0.205	0.116	0.055	0.137	0.072	0.007	0.024	0.017	0.041
2016	0.032	0.114	0.061	0.121	0.207	0.132	0.036	0.118	0.082	0.014	0.011	0.014	0.057
2017	0.056	0.130	0.053	0.109	0.258	0.096	0.053	0.093	0.065	0.031	0.022	0.003	0.031
2018	0.065	0.129	0.056	0.071	0.247	0.100	0.021	0.153	0.071	0.009	0.026	0.000	0.053
2019	0.053	0.114	0.045	0.122	0.216	0.139	0.020	0.118	0.078	0.020	0.045	0.004	0.024

Table G.11. Distribution of Home Runs, L-L

	1	2	3	4	5	6	7	8	9	11	12	13	14
2008	0.007	0.018	0.011	0.008	0.019	0.014	0.004	0.012	0.013	0.002	0.003	0.000	0.001
2009	0.009	0.018	0.012	0.009	0.021	0.012	0.003	0.013	0.013	0.002	0.002	0.001	0.002
2010	0.008	0.017	0.007	0.010	0.016	0.017	0.005	0.012	0.009	0.002	0.001	0.000	0.002
2011	0.009	0.013	0.008	0.008	0.017	0.013	0.002	0.011	0.012	0.002	0.001	0.000	0.001
2012	0.009	0.015	0.011	0.010	0.018	0.012	0.004	0.009	0.012	0.001	0.002	0.000	0.001
2013	0.005	0.015	0.005	0.008	0.014	0.019	0.003	0.010	0.009	0.001	0.002	0.000	0.003
2014	0.005	0.007	0.005	0.007	0.014	0.012	0.004	0.011	0.010	0.000	0.002	0.000	0.003
2015	0.008	0.011	0.009	0.009	0.015	0.013	0.004	0.012	0.010	0.000	0.001	0.000	0.003
2016	0.005	0.014	0.010	0.009	0.016	0.016	0.003	0.010	0.012	0.001	0.001	0.000	0.004
2017	0.009	0.016	0.009	0.010	0.023	0.012	0.005	0.010	0.011	0.002	0.001	0.000	0.002
2018	0.010	0.016	0.009	0.006	0.020	0.012	0.002	0.015	0.012	0.001	0.002	0.000	0.004
2019	0.011	0.019	0.010	0.014	0.023	0.023	0.002	0.015	0.015	0.002	0.004	0.000	0.002

Table G.12. Proportion of Home Runs, L-L

Appendix H: Control Charts for Pitcher-Batter Handedness Splits

Given the large amount of control charts, we put them in a separate document called *AppendixH_HomeRuns_2019.pdf*. We also eliminated the control charts on the distribution of home runs. In order to make the file more stand-alone, we have included combined graphs before each set of control charts. This creates duplication of graphs from Appendix G. The order of the graphs in the file is as follows, for each of the R-R, R-L, L-R, and L-L pitcher-batter handedness match-ups:

- 3 line graphs of proportions of pitches in zone: distribution of pitches, distribution of home runs, proportion of home runs
- 12 Control charts for distribution of pitches
- 12 Control charts for proportion of home runs

This gives a total of (3+13+13)*4 = 116 graphs.