

Analysis of Amateur Baseball Player Athletic and Visual Tests

Simi Bleznak^{1*}, Maxwell Brown^{1*}, Julia Choi^{1*}, William E. Carson IV^{2*}, Hunter Klein^{2*}, Joel Huber^{3*}, Greg Appelbaum^{3*}

¹Undergraduate Data Scientist (equal contribution), ²Project Manager, ³Project Lead, *Duke University, Durham, NC 27708



Background

Exploring how visual, cognitive, and physical abilities relate to physical performance can provide insight into the development of athletes. Strong relationships between these variables could influence scouting efforts and training regiments under the notion of predicting and improving in-game athletic performance.

Goals

- Explore *inter-data* relationships between different datasets.
- Determine whether test performance can predict position.
- Determine how test performance varies with age and repetition.
- Compare test performance on visual assessments relates to performance on physical assessments.

Predicting Position from Player Tests

We created several logistic regression models that identified the most important variables for predicting position.

- Every predictor has an associated Bayesian Information Criterion (BIC) value. If addition of a predictor to the model results in a lower BIC, it is added to the model as a predictor.
- Each positional group model started with all PDP and RightEye metrics as predictors and through backwards selection, predictors were removed that did not lower the model's BIC.

Table 1: Model predictors for each positional group model and corresponding area under the receiver operating characteristic curve (AUC) obtained during model prediction on hold-out test set. Significant predictors ($p < .05$) corrected for multiple comparisons are bolded and italicized.

Position	Predictor 1 (coef.)	Pred. 2 (coef.)	Pred. 3 (coef.)	Pred. 4 (coef.)	Pred. 5 (coef.)	AUC
Catcher	30 yd Dash (.227)	Height (-.034)	Weight (.003)	–	–	0.62
Middle Infielder	30 yd Dash (-.353)	Weight (-.006)	–	–	–	0.78
Corner Infielder	Broad Jump (-.005)	Weight (.004)	–	–	–	0.75
Outfielder	Ground Contact Time (-.463)	30 yd dash (-.253)	Counter Movement Jump Height (.030)	–	–	0.60
Pitcher (LH)	30 yd Dash (.155)	Height (.011)	Leg Strength Ratio (.132)	–	–	0.69
Pitcher (RH)	Height (.042)	Counter Movement Jump Height (-.017)	RH grip strength (-.003)	Sports Function Score (.004)	Sports Mechanics Score (-.004)	0.73

Athletic and Visual Tests Data

We explored two datasets which measured amateur baseball player's athletic (PDP) and cognitive abilities (RightEye).

USA Player Development Pipeline (PDP) Dataset:

- Measures players' speed (30-yard dash), lower body explosiveness (broad jump), and grip strength.
- Various tests that measure players' visual and cognitive skills.

RightEye Dataset:

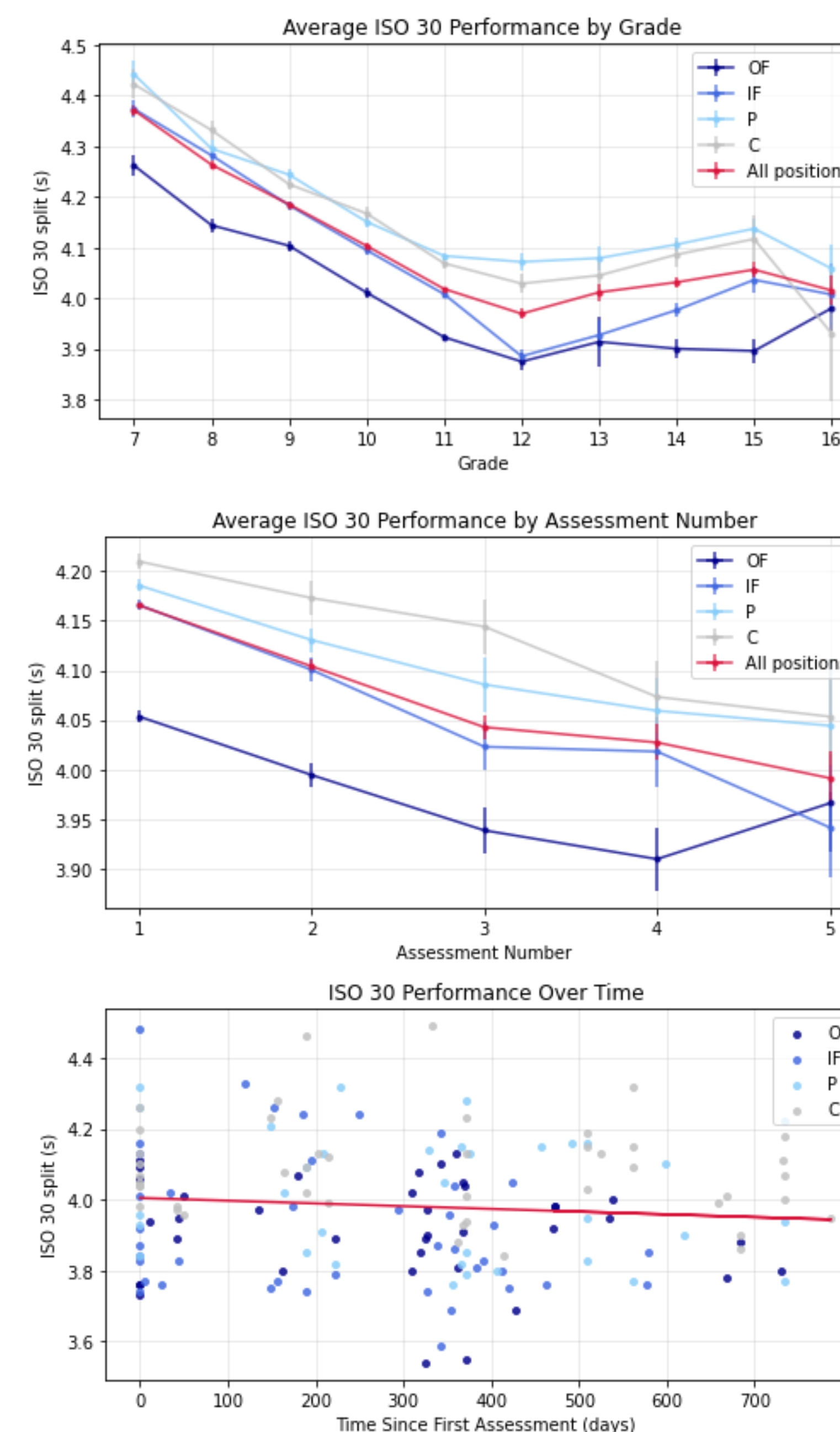
- Consists of several eye-tracking and visual-motor test scores that evaluate reaction time and concentration proficiency.



Figure 1: Examples of PDP data collection (left and middle) and RightEye data collection (right).

Temporal Trends in Player Tests

Figure 2: Average physical performance improves over younger age groups and declines over college age groups, but visual assessment scores improve steadily over all grades (top). Average performance improves as players repeat the same assessments at different events (middle). Player performance improves over time on an individual basis (bottom).



Cross-Sectional Prediction of Tests

We used linear regression to explore whether performance on certain PDP tests can be used to predict RightEye cognitive scores.

These PDP tests were selected due to the combination of physical and mental skills they test. Low R^2 value of the best fit line between true and predicted scores indicated little relationship between the two datasets. However, regression coefficients show expected relationships between PDP tests and RightEye brain measures.

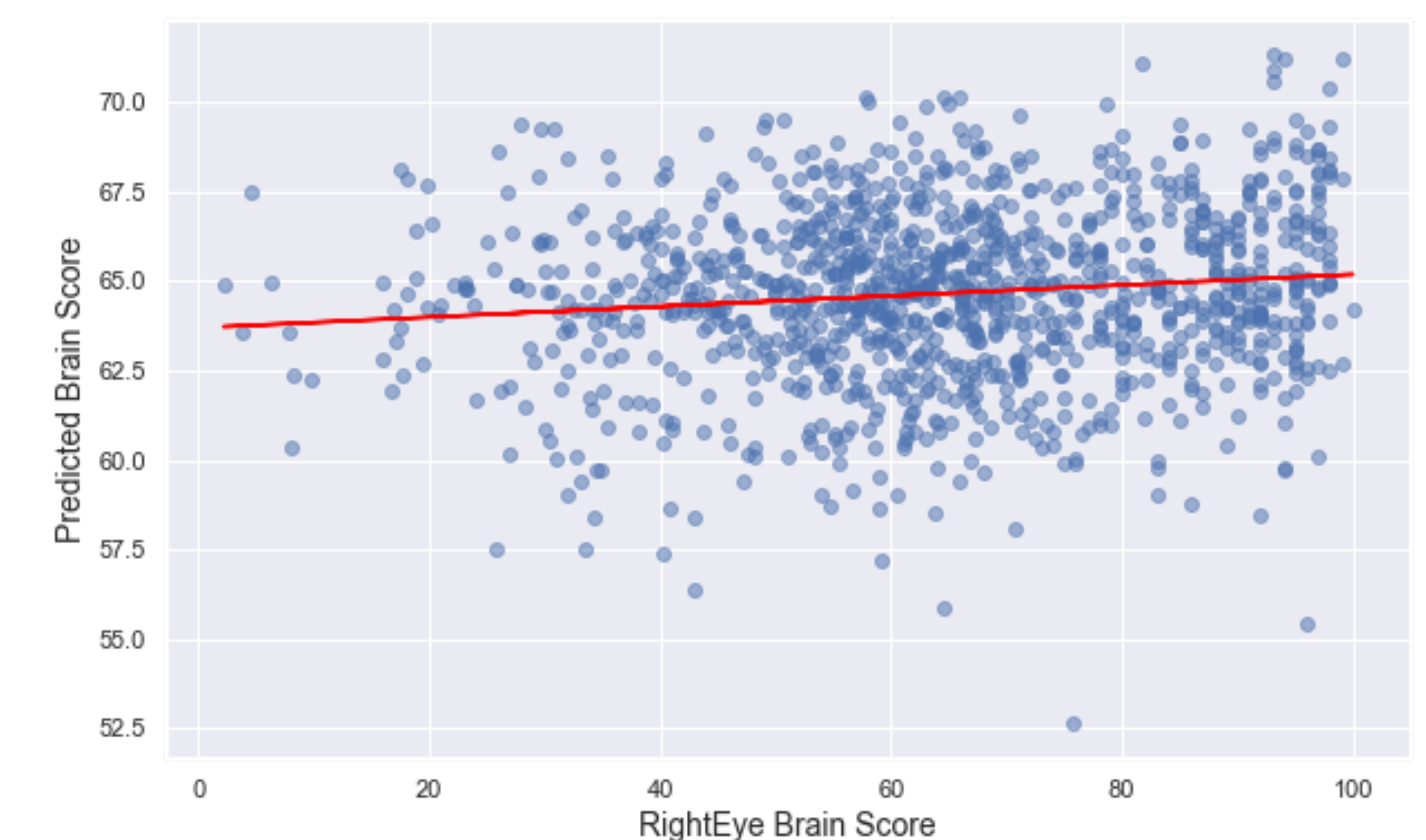


Figure 3: True vs predicted RightEye Overall Brain Score values and best fit line ($R^2 = 0.014$) as a function of PDP tests reaction time (coef. = -0.613), Green Box test (coef. = -0.642), Green 3 test (coef. = -1.940), and Hawkeye test (coef. = 0.276).

Conclusions

We used linear regression analyses, logistic regression models, and longitudinal analysis to understand how players' physical and cognitive tests change over time and to determine the predictive power of this data. Our analyses show that

1. Certain PDP and RightEye measures demonstrate predictive power of position.
2. Player athletic, cognitive and visual tests improve over time and through repetition.
3. Little relation exists between RightEye and PDP data.

Acknowledgements

We thank USA Baseball for their collaboration and for providing multiple datasets that made this work possible.