All data used in our analysis is contained in the data folder or can be acquired by running the import\_and\_save\_data.ipynb notebook.

Note the Rangers provided the proprietary ground ball and fly ball out data. We were not permitted to share that data, so to meet the open-source requirement from Sloan, we transformed that data to anonymize it. We applied a randomly generated affine transformation to each column in those datasets. The last column in those datasets is the response variable: the is\_out column. Thus, it is possible to fit equivalent logistic regression models to the Rangers proprietary ones, meaning that you can fit a model that makes the same prediction on each row of data that the Rangers model makes, but the coefficients of the model will not match the Rangers proprietary model and the values in the columns do not match the proprietary data because of the affine transformations. This allows us to maintain the proprietary nature of the data while publicizing it. If you have any further questions, feel free to email me at [wmmelville@gmail.com](mailto:wmmelville@gmail.com).

The great majority of our analysis was done using the public data acquired from the import\_and\_save\_data.ipynb notebook anyways. The only results that can’t be replicated using public data are the Joey Gallo bunt results and the last paragraph of section 3.5. Everything else was done using public data.