**AI for Baseball - Summer Plan**

Broad project outline:

Create heat maps of batting statistics (e.g. hit probability, home run probability) and quantify variation over various variables (e.g. hitter, pitcher, ballpark, seasonal). Compare different methods for constructing heat maps - generalized additive models, neural networks, and gaussian process models.

If time, use these heat maps to understand how changes or differences in plate approaches would impact player or team performance.

Schedule:

|  |  |  |
| --- | --- | --- |
| **Date** | **In-club activities** | **Optional out of club activities** |
| 16th May | - Meet group & introductions  - Discover shared interests  - Brainstorm ideas & look at data | - look for relevant papers / tutorials / continue exploring data |
| 30th May | - Loosely define research idea  - begin exploratory data analysis (EDA) | - EDA |
| 13th June | - EDA  - Discussion of modeling approaches & start fitting models | - EDA  - Modeling approach research |
| 27th June | - Fitting models | - EDA  - Work on modeling code |
| 11th July | - Fitting models | - Work on modeling code |
| 25th July | - Fitting models | - Work on modeling code |
| 8th August | - Fitting models | - Work on modeling code |
| 22nd August | - Wrap up  - Discuss what is left to be done  - Ensure github is up to date | Summer onwards  - continue data analysis  - manuscript preparation  - conference attendance |

Resources to get started:

|  |  |
| --- | --- |
| <https://billpetti.github.io/baseballr/> | R package for scraping baseball data & functions to calculate common metrics |
| <https://bayesball.github.io/shinybaseball/> | Shiny app to explore different types of baseball data and analyses |
| <https://bayesball.github.io/BLOG/> | Blog posts from prominent baseball researcher |
| [Analysing Baseball Data with R](https://drive.google.com/drive/u/0/folders/1FH-IDmz565BX36Xppagzpr84an2HZGpb) | Helpful book on baseball statistics in R |
| <https://baseballwithr.wordpress.com/> | Wordpress blog accompanying above book |
| <https://baseballwithr.wordpress.com/2016/09/19/constructing-heat-maps-for-avg/> | Simple version of the analysis we will run |
| <https://baseballsavant.mlb.com/statcast_search> | Website for downloading batter or pitcher specific data |

Groups:

* Modelling hit probability
* Modelling home run probability
* Game simulation
* Open group (different questions or methods)