Sam Graf

Final Project Proposal

Title: Analyzing Division 3 baseball data with R: what metrics lead to success?

My dataset is a collection of all D3 pitching and hitting data from last year. With this data, I plan to investigate the differences between winning and losing teams using different statistics. In turn, this helps to answer the question “what metrics can be used to help predict team success?”

Additionally, I will also look at individual statistics to see if there may be a better way in identifying All-Americans using non-traditional metrics.

Datsets can be found here, available for csv download.

<https://sabermetrics.cs.illinois.edu/pitching/ncaa_baseball/>

<https://sabermetrics.cs.illinois.edu/batting/ncaa_baseball/>

What sort of offensive stats best predict offensive performance? Pitching stats? What about team performance?

All americans: did they get them right? Can cluster by all-americans and by non all americans and see what kind of stats they’ve got in common between them and if that agrees with some other metrics used.

What are the differences between the number of x stat for example between winning and losing teams? In order to get at what stats are most important for predicting winning performance