Final Project Proposal

Xiangchen Guo

Due on April 25 at 11:59 pm

Your final project proposal is meant to ensure that you and your final project group choose a project that is rigorous, interesting, and can be completed in time. Final projects are due on 05/15 at 11:59 pm, and you are allowed and encouraged to work in groups of 2-3. Completion of this project proposal is worth 50 points.

Recall that there are 375 points allocated towards your final project. As just mentioned, this proposal is worth 50 points. There are 75 points allocated for your presentation. This presentation will be in the form of a 5-10 minute recorded presentation. The remaining 250 points is allocated to project materials and deliverables. Your presentation and your project materials are all due on or before 05/15 at 11:59 pm.

Project Rubric (or Lack Thereof): There is no formal rubric for your final project. This is intentional: a rigid rubric would be too limiting, as the most compelling projects come in many forms. Some may be large-scale Shiny apps rich in functionality. Others might be small but polished apps with a clever, focused idea. Still others could take the form of detailed written reports offering deep analytical insight.

The only guiding principle is this: think of something interesting to create, and then create it as if it were meant to be used or read by an interested stakeholder. As Dr. Arnold O. Beckman once said:

"There is no satisfactory substitute for excellence."

Instructions. Fill out this form.

Group members:

Xiangchen Guo Zijian Shen Sean Pang

Project description. Provide at least one paragraph describing what your project is about:

This project will involve examining Major League Baseball (MLB) performance data that can be used through an interactive R Shiny webapp. The performance metrics will come from Statcast data taken from Baseball Savant and the Lahman baseball database. Our intention is to examine and visualize performance metrics for hitters and pitchers while focusing on the trends in player performance (e.g., exit velocity, launch angles, pitch type, and strike zone efficiency) over the last several seasons. The app will also help users explore how performance metrics/app metrics influence the outcome of games, as users will also be able to compare across players and teams. This project explores the intersectionality of statistical insights and pleasantries of visualizations, providing useful ways for baseball enthusiasts and data science learners to explore data.

Project deliverables. Provide at least one paragraph describing what you plan to produce:

We will develop an R Shiny app that will have multiple dashboards for visualization of pitch-level information (e.g., strike zones heat maps), player-level information (e.g., by season average exit velocity), and team-level comparisons. Users will have the ability to filter by team/year, position, player, and navigate dynamic plots and summarized tables of values. We will add a brief write-up explaining what methodology was done in R Markdown in the app to demonstrate how the data was cleaned, combined, and visualized. All code, documentation, and deployed app will be at GitHub with a link put in the report upon completion.

Project submission. Provide at least one paragraph describing how you plan to submit your final project materials:

Final submissions are to be hosted in a collaborative GitHub repository. These include the entire R Shiny program, CSV data files (or download script and preprocessing scripts), plots, and an R Markdown document of our workflow of data preparation and analysis. Setup details are documented in a README, and we host the live Shiny program on shinyapps.io. We are also required to publish a short video tour of significant aspects of the program.

R shiny layout. Provide a sketch of the proposed layout of your Shiny app (ignore this if you are not creating a Shiny app):

The layout of our Shiny app will contain four tabs:

Player Statistics – Select a player along with a season to compare batting/pitching statistics over years through line plots, radar charts.

Heatmaps - Density plots of where a pitcher throws balls of different types (fastballs vs. sliders, i.e.)

Team Comparison – Bar charts and tables comparing teams by season based on OPS, ERA, WAR, etc.

Historical Explorer – Timeline viewer of notable games or events from Lahman dataset

Each page shall contain interactive filters (selectInput, sliderInput) as well as reactive graphical outputs (plotlyOutput, tableOutput). Side panels shall regulate user input, and main panels shall show outcome interactively.