Caroline Thomas

A Novel Method for Offensive Data Collection and Performance Analysis for Colorado State University Softball

Goals and Motivation

The primary goal of my thesis project was to develop a streamlined method for data collection for the Colorado State University Softball team to use during practice and game situations while on offense. Currently, the data collection method employed by the team is primarily through the program Right View Pro during home games and limited paper-and-pencil data collection during practices. The Right View Pro software consists of video recordings of the games which are tagged by a manager to populate a spreadsheet containing play-by-play data. While still useful, this method is error-prone and additionally, does not directly incorporate the player's assignment ("game plan") as given by the coach. This includes personalized player information about the locations of pitches that players should or should not be swinging at which is necessary to assess player execution of their given "game plan." This is a quality that is of primary interest to the coaches in order to structure practices and make decisions about the starting lineup. Hence, the coaches make limited use of Right View Pro data, and resort to paper and pencil data collection techniques. During practices, game-like situations are simulated, however, data collection of these scenarios is limited to what can be done by hand, so this data is sparse and seldom used beyond that day's practice. In recent years, the world of sports has transitioned into a data-driven industry, with most professional organizations and many university teams relying on advanced analytics. While our softball team here at Colorado State University continues to climb in national ranking, improving their collection and use of data would allow coaches to more effectively coach their players.

Once a more efficient data collection method has been created, this data can be used to assess how well CSU softball players execute their given game plan. That is, not whether or not they got a hit, but rather how well they were able to execute their coach's instructions i.e swing at pitches in the zone where they perform the best, and lay off of pitches that they are not as good at connecting with. The coaches are interested in having players who are coachable and able to correctly execute a given plan of action so an assessment or "grade" indicating how well players are doing this would be helpful for the coaches when determining their starting lineup as well as determining what should be focused on in practices. These "grades" could also serve as motivation for the players, highlighting what they can improve upon and provide an accurate assessment of their performance from the coaches' perspectives, even if their game statistics (batting average, runs scored, etc.) do not seem as good. To achieve these goals, the data collected needs to include information about the situation (count, runners on base, etc.), player-specific information about which pitches they should be looking for, and finally pitch and outcome information about what pitch was thrown, where it was in the strike zone, and the outcome of the at-bat. Outcomes can then be classified into wins and losses according to the given situation and instructions from the coach which can be directly linked to a "grade" for the player or team as a whole.

My reason for choosing this thesis project is that it is a fusion of two of my passions-- that is, my love for sports and my love for statistics and data analytics. I am a Statistics major so my interest in data analytics speaks for itself, but I am always blown away by the power of data and the tools that have been created to process it. Data analytics opens so many doors and can answer so many questions when the correct data is collected and used resourcefully. My love for sports stems from a childhood of playing almost every sport. Softball, though, was the one sport that really stuck with me, resulting in 12 years of playing competitive softball. Since retiring from the game after my senior high school season, I have been looking for an opportunity to find my way back to softball, even from the sidelines. This past

summer, I had a sports statistics internship working with the CSU Women's Volleyball team in conjunction with the Statistics Department, which I absolutely loved. I already knew that I loved sports and statistics but my work over the summer really showed me that sports analytics is something that I am really passionate about and would possibly like to go into as a career in the future or at least pursue as a hobby. Prior to beginning this project, I reached out to the CSU Softball coaches to ask if they would be interested in working on a project together and the head coach, Jen Fisher, was very interested. After many meetings with Coach Fisher, it was clear that she was interested primarily in assessing the coachability of players and how well they are able to execute the instructions she gives them, which led into the current project. Initially, I wanted to focus on more of the data analysis with the Right View Pro data that had already been collected, however, flaws in these data and a lack of player-specific information prompted me to design my own method for data collection that would allow me to have all the information I need for further analyses.

Methods

This project first began by meeting with the Colorado State University Softball Head Coach, Jen Fischer to discuss possible projects to work on. After much brainstorming, we arrived at the conclusion that finding a way to grade players based on how well they executed Coach Fisher's instructions was of peak interest to the coaching staff. This would allow Coach Fisher to evaluate how coachable players are and to demonstrate to players that following her instructions leads to good outcomes, even if individual game statistics are not as high as desired. Coach Fisher detailed the team's current methods of data collection and how they use that data, which began to expose some areas for improvement. We determined that in order to appropriately answer the

questions that Coach Fisher was asking, we would need to collect different data than what was already accessible, which prompted the idea of this data collection application.

Once this project was decided on, the first step was to conduct a brief literature review to see if any similar work had been done. To no surprise, there was no record of any project like this to be found. The primary component of my project was to create a working R Shiny app, using the computer program R. While I was very familiar with R at the onset of this project, I had never touched R Shiny before. R Shiny functions very differently than base R or a typical R Studio script, so I had to spend quite a bit of time familiarizing myself with the structure of R Shiny and how to build different components of an app. I did this through online tutorials through the R Studio website, YouTube, an unpublished book called *Mastering Shiny* by Hadley Wickham. After drafting an overview of what components should be included in my app, I slowly added the pieces to my app as I followed along with these tutorials and if necessary, I resorted to Google and online forums such as Stack Overflow and R Bloggers to answer any further questions. Additionally, I met with Ben Prytherch and Brooke Anderson to aid with debugging my code. This portion of my project involved a lot of trial-and-error work as I tested out various lines of code and altered them until I reached the desired output.

After completing a first draft of the app, my goal was to meet with Coach Fisher again to get her feedback on the functionality of the app and the ease of use. Using her feedback, I could refine the app to improve its efficiency and usefulness for the coaching staff. Additionally, I planned to test the app at a real practice, to evaluate its intended functionality. However, due to the COVID-19 pandemic, these steps were unable to be completed, but may be carried out in the future once softball practices and meetings are permitted again.

The next step in completing my project was to take the file that the software application outputs and use it to do further data analyses. This involved writing a script using R Markdown that would perform a basic summary of the data and produce concise plots that the coaches could easily interpret and make use of. This script was tested using mock practice data that I created, and revisions were made to improve it and to ensure that a coach or manager with limited analytics knowledge could still use it.

Final Products

My final product consists of several items which were submitted to my thesis advisor and to the Honors Office. Additionally, these materials have been posted to a GitHub Repository found at: https://github.com/thomascr9/Honors-Thesis.

(1) **Shiny App:** The first deliverable is a working online Shiny App that is compatible with laptops and tablets. This app is hosted through the online server, shinyapps.io, and can be used for data collection during softball practices, specifically, the batting component of practice. The app collects information about player name, current situation (number of balls and strikes), personalized player "goals" (i.e. which pitches the player is hunting for), as well as information regarding the outcome of the at-bat, the pitch location, and whether or not the pitch is denoted as a win or a loss for the batter. The app has an easy-to-use interface that makes data collection efficient for anybody inputting the data. Further, the app displays the data that has been collected and has the capability to filter this data to make it easy to reference previous pitches quickly. The app can be run through R Studio or can be accessed online through this url: https://caroline-thomas.shinyapps.io/softballapp/.

(2) **Example data:** The data from a given practice can be downloaded as a comma separated values (.csv) file which can then be analyzed further in R or in other programs such as Excel. An example of the resulting spreadsheet was created.

(3) **Software tutorial:** I have also made a tutorial containing two screen recordings on how to use the app for team coaches and managers who might be using my application. There is one tutorial for how to use the app and another on how to run the basic report R Markdown file in R Studio. I have created a mock practice design for this tutorial to simulate what an actual practice might look like when using the app. These tutorial videos have been posted on YouTube and can be accessed with the link:

https://www.youtube.com/playlist?list=PLonGs-oj4UGBcBxyflNtDNxNG9d725Okb

(4) **Hitting report:** Additionally, I have also written an R Markdown script that performs basic analyses on these data after the practice has ended. This R script is organized and commented on in a manner that is easy to understand and use for others with limited experience using R. This script includes an output of summary statistics and simple plots that provide an easy way to get a basic report on how players performed during practice so that coaches can make quick adjustments the next day. Again, this script was created using the mock practice data that I created.

(5) **Reflection paper:** Finally, I have written this reflection paper to explain and reflect on my thesis experience and detail the project that I have done.

Contribution to Learning and Future Career

This project contributed to my learning experience and prepared me for a career in several ways. There are several technical skills that I either learned or improved throughout this semester

of working on my thesis. The primary one that stands out to me is my R programming skills greatly improved, especially in working with R Shiny, which is an element of R that I had no previous experience with. Shiny apps are becoming very popular due to their easy-to-use interface and the complexity of plotting and data analysis that can be done with them while still maintaining their ease and appearance. Knowing how to create Shiny apps is a useful skill in any future job I may have, and in fact, I have had several of my peers go on to have jobs in statistics that involve building these apps. Additionally, I was able to improve my data analysis skills by thinking about which statistical methods should be used in the content of my specific goal and the data I have, and how to perform the necessary computations in R. Along this note, any career in statistics will involve spending a large portion of my time collecting, cleaning, and manipulating data before I can create a model or perform an analysis. My thesis relied very heavily on proper data collection and data format, which gave me a deeper insight into how data should be collected and organized when conducting a project such as this.

Further, there are several broader lessons that I learned throughout working on my thesis. First, my statistical communication skills greatly improved. My ability to convey data analyses through summaries and plots in a way that is clear and concise to someone with limited statistical knowledge is a skill that will be critical in my future. The coaches and managers of the softball team do not have a deep knowledge of statistics so one of my primary goals was to create an app and an analysis output that would be useful to them, given the knowledge that they have. The most advanced coding and statistical methods are still meaningless if they cannot be conveyed in an understandable manner so this skill, in particular, was a big takeaway for me. Another important skill that I was introduced to was how to start with a problem to solve rather than with data and a predetermined analysis method. In the academic setting, assignments typically start

with a pre-cleaned data set and a specific statistical method that is supposed to be implemented. However, in the real-world a company might ask you to solve a problem or answer a question but you are not given data or the steps to complete the project—instead, you must problem solve to figure out what data you need, how to access said data, and then finally the best approach to using the data. The process I used to complete my thesis was much more similar to this real-world scenario which was helpful in teaching me the difficulties and intricacies that come along with the types of problems I will be faced with in a future career.

Overall Reflection of Experience

Completion of my thesis this semester was riddled with troubleshooting computer issues and debugging code. Using R Shiny, which was an aspect of R that was completely unfamiliar to me, had a steep learning curve and finding adequate resources to aid me was difficult. The majority of my time was spent employing trial-and-error methods to create my app which left little time for more interesting analysis and more sophisticated design elements. Following the end of the semester, I am planning on continuing to pursue this project in order to fine-tine the details and make it as useful as possible for the softball team. Despite my challenges, I have found this project to be incredibly rewarding in that I have learned a new skill that will serve me well in my future endeavors and I have experienced first-hand the obstacles that come with designing and debugging an app, learning a new program, and doing all this in the midst of a global crisis.

There are some aspects of my thesis that I could have done better, first and foremost being time management and proactive work. I tended to try to solve problems on my own for much longer than I probably should have before reaching out to my thesis adviser or other faculty members to get their help. While I think this gave me the opportunity to solve problems on my

own, I think I could have saved time and frustration had I gotten help sooner. Additionally, I believe I could have done my work more consistently on a daily basis rather than do it in larger time intervals less frequently.

Overall, my thesis experience has been both frustrating and fun at times. I have felt both proud and independent in my work as well as lost and on my own. The juxtaposition of these feelings has made for a love-hate relationship with my project, but all-in-all, I have enjoyed the experience and feel like I have gained many useful skills throughout the duration of the project.