Disclaimer: Obviously, I have not yet completed my project so more details will be added, or changes will be made, to their respective sections once the project has been completed.

**Papers must include:**

**(a) an explanation of the topic area and reasons for choosing it**

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 provide personalized player information about which pitches players should or should not be swinging at which makes it difficult to assess player execution of their given “game plan” which is a quality that is of interest to the coaches. Hence, the coaches make limited use of this data, despite its potential. Further, 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. 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 does incredibly well, improving their collection and use of data could have many positive effects on their player performances and outcomes. Thus, improving the methods of data collection so that the data being collected can be used to assess player performance in a more effective manner is of importance to the coaches.

Once a more efficient data collection method has been created, this data should 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. In doing this, these “grades” would also serve as motivation for the players to show them what they can improve upon and how well they did in the eyes of the coaches, even if their game statistics (batting average, runs scored, etc.) don’t seem as good. To achieve this, 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 or laying off of, 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. We classified outcomes into wins and losses according to the given situation and instructions from the coach, in order to easily calculate a proportion of wins and losses 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 this project 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 kind of 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.

**(b) the steps taken to complete the project**

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 measure or “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 show the players themselves that following her instructions leads to good outcomes in the end, 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 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, and 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 to answer any further questions. 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, we met with Coach Fisher again to get her feedback on the functionality of the app and the ease of use. Using her feedback, revisions were made to the app to improve its efficiency and usefulness for the coaching staff. Additionally, the app was tested at a real practice, again, to evaluate its intended functionality.

The next step in completing my project was to take the comma separated values (.csv) file that the 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 and revisions were made to improve it and to ensure that a coach or manager with limited R knowledge could still use it.

**(c) a description of the final work**

My final product consists of several items. The first is a working online Shiny App that is compatible with laptops and tablets. This app is hosted through the online server, shinyapps.io, and is 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 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 this resulting spreadsheet has been printed out and submitted to the Honors Office. I have also made a tutorial containing screen recordings and written instructions on how to use the app for team coaches and managers who might be using my application.

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 coaches or managers 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. Finally, I have written this reflection paper to explain and reflect on my thesis experience and detail the project that I have done.

**(d) an assessment of the importance of the creative activity and how it contributed to the learning process and/or prepared you for a 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 mangers 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 that would still be useful for them and easy for them to perform, 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.

**(e) a brief reflection about the overall thesis experience**