1. A Novel Method for Offensive Data Collection and Performance Analysis for Colorado State University Softball
2. This project is a fusion of two of my passions-- that is, my love for sports and my love for statistics/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 which is awesome! Statistics also involves computer programming which is something that I have grown to enjoy over the past couple years. 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 like to go into as a career in the future. Near the end of the summer, we reached out to the CSU Softball coaches to ask if they would be interested in working on some similar project with us, going forward and the head coach, Jen Fisher, was very interested!
3. The primary goal of my thesis is to develop a streamlined data collection method 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. This consists of video recordings of the games which are tagged by a manager to populate a spreadsheet containing play-by-play data. This method is error-prone and additionally, does not provide personalized player information about which pitches they should or should not be swinging at which makes it difficult to assess player execution of their given “game plan” which 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 are done via pen-and-paper so their use is limited to what can be done by hand. I would like to improve the methods of data collection so that the data being collected can be used to assess player performance in a more effective manner that is of interest to the coaches. Once a more efficient data collection method has been created, I would further like to begin assessing how well CSU softball players execute their given game plan. That is, not whether or not they got a hit, but rather how well were they 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 do their job so they would like a way to “grade” how well each player is doing this. 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, ERA, etc.) don’t seem as good. To achieve this, the data collected will need 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 information about what pitch was thrown and where it was in the strike zone.
4. My first step will be completing a literature review of current methods for evaluating player performance in softball and baseball in order to investigate what data should be collected in order to achieve my goals. Softball analytics is a largely unexplored field, so I am not expecting great results from this. Once I have an idea of what data needs to be collected, I will use the computer programming software R to create a Shiny App that will be compatible with laptops and tablets for easy data collection during practices and games. This application will collect data at the touch of a button and will populate a comma separated values (.csv) file with the data that can later be used to do data analysis on player performance. Finally, I will prepare a script that will perform basic data analysis with the input of this csv file.
5. Key Resources Needed:   
   1. “Learn Shiny.”RStudio Inc., 2017. https://shiny.rstudio.com/tutorial/
   2. Bailey Fosdick, Assistant Professor, Department of Statistics
   3. Ben Prytherch, Instructor, Department of Statistics
   4. Jen Fischer, Head Colorado State University Softball Coach
   5. Marchi, Max, Albert, Jim, Baumer, Benjamin S. *Analyzing Baseball Data with R.* 2nd ed., CRC Press, 2019.
   6. Journal of Open Source Software (JOSS)
   7. Journal of Quantitative Analysis in Sports (JQAS)
   8. Petti, Bill. “Baseball Tools.” Github, 2020. https://billpetti.github.io/baseball\_tools\_home/
   9. Wickham, Hadley. *Mastering Shiny.* O’Reilly Media, 2020.
6. Timeline to Completion:  
     
   Formal Proposal: 2/3/2020  
   Literature Review: 2/17/2020  
   Mock App/Spreadsheet Design: 2/17/2020  
   Basic App (Version 1): 3/9/2020  
   App Revisions: 3/23/2020  
   Basic Script for Data Analysis: 4/6/2020  
   Penultimate Draft: 4/17/2020  
   Oral Presentation: 5/4/2020  
   Final Copy: 5/11/2020  
   Thesis Evaluation Form: 5/15/2020
7. The final product will consist of several items. The first will be a working online Shiny App that is compatible with laptops and tablets. This app will be used for data collection during softball games and practices. It will have both a game mode and a practice mode and will include information about pitch type, game situation (or simulated situation), and personalized player “goals”. This app will populate a csv spreadsheet with the data to be used in further analyses. I will submit a tutorial containing screen recordings and written instructions on how to use the app as well as a paper copy of the resulting spreadsheet. Second, I will prepare and submit an R Markdown document (paper copy), that will perform basic analyses on these data after the game or practice has ended. This R script will be organized in a manner that is easy to understand and use for coaches or managers with limited experience using R. Finally, I will submit a 7-page written reflection paper detailing my thesis experience including explanations of my interest in the subject, methods and steps taken to complete the project, the final product, the importance of the project, a reflection of the overall experience, and appropriate citations of resources used.