Laurel Ayuyao

Professor Barron

ITAO 40520

22 October 2020

Project Proposal

# Motivation

In the world of sports analytics, volleyball has been one of the last sports to try to use data to drive strategy and to try to improve team and player performance. There is still very little data from volleyball matches that are publicly available. However, as evidenced by the increased adoption of analytics in other sports, including baseball and football, it would most likely be beneficial to use data analytics in volleyball in order to drive decision-making and to derive insights that can be used to better player performance.

Since there is barely any previous work in the volleyball space, I am looking to explore many different aspects of sports analytics. Some of the problems that I would like to look into is whether players make more or less errors near the end of sets, how much calling a timeout affects whether a server will make their serve, and which areas of the court should a hit be directed at to be more effective, among others. Once I have a chance to explore the data more deeply, I am certain that there will be additional problems that will be interesting to examine. With the limited amount of data, this type of analysis would probably be most beneficial to a coach or player who is looking to use data to make decisions about their strategy.

# Problem Framing

In order to answer the aforementioned questions, I am planning to use play-by-play data to get a detailed analysis of performance. I will most likely use both statistical calculations and graphical representations to try to draw insights. I plan to explore a variety of factors that might have an effect of player performance and game outcome and dive deeper as the project progresses.

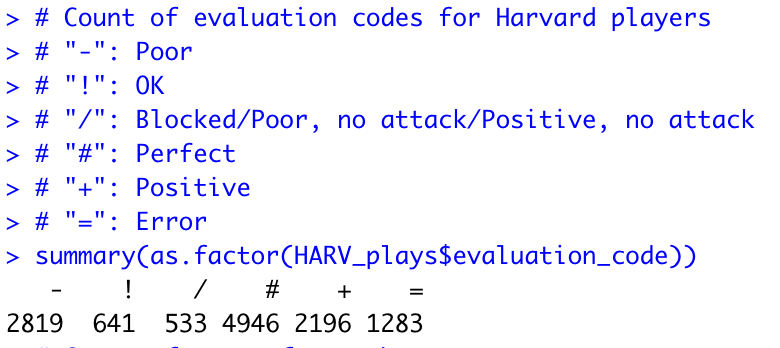
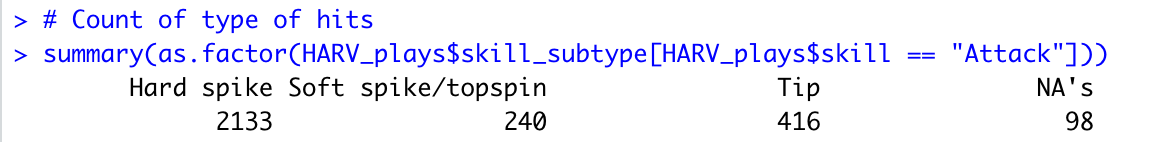
# Data Overview

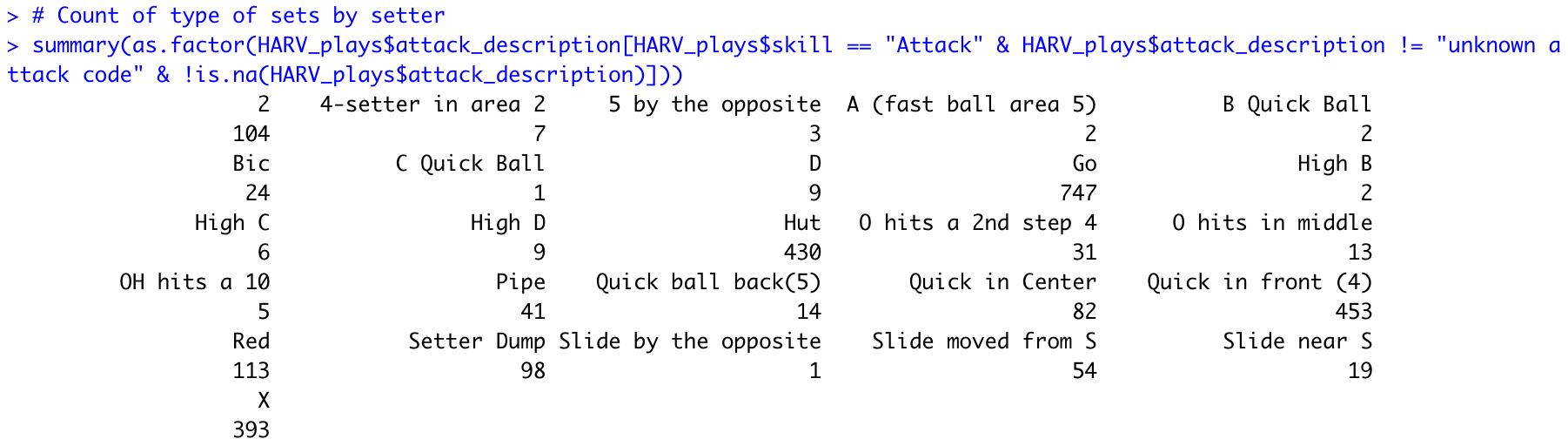
I found a dataset that has all the play-by-play data for Harvard University's 2018 season which was recorded using a software called Data Volley. A user macfields uploaded this data to a [Github repository](https://github.com/macfields/mfields_finalproject) for their own project, but most of their analysis was done using a Shiny Dashboard, which I am unfamiliar with. To import and parse my data, I will be using a package called [Datavolley](https://github.com/raymondben/datavolley) which was created by Github user raymondben. The data can be downloaded from Github, but I have also imported and [shared it using Google Drive](https://drive.google.com/file/d/1snF1GFCm_cEa5VJX0ukYeWBiLAIPZjVQ/view?usp=sharing) for convenience.

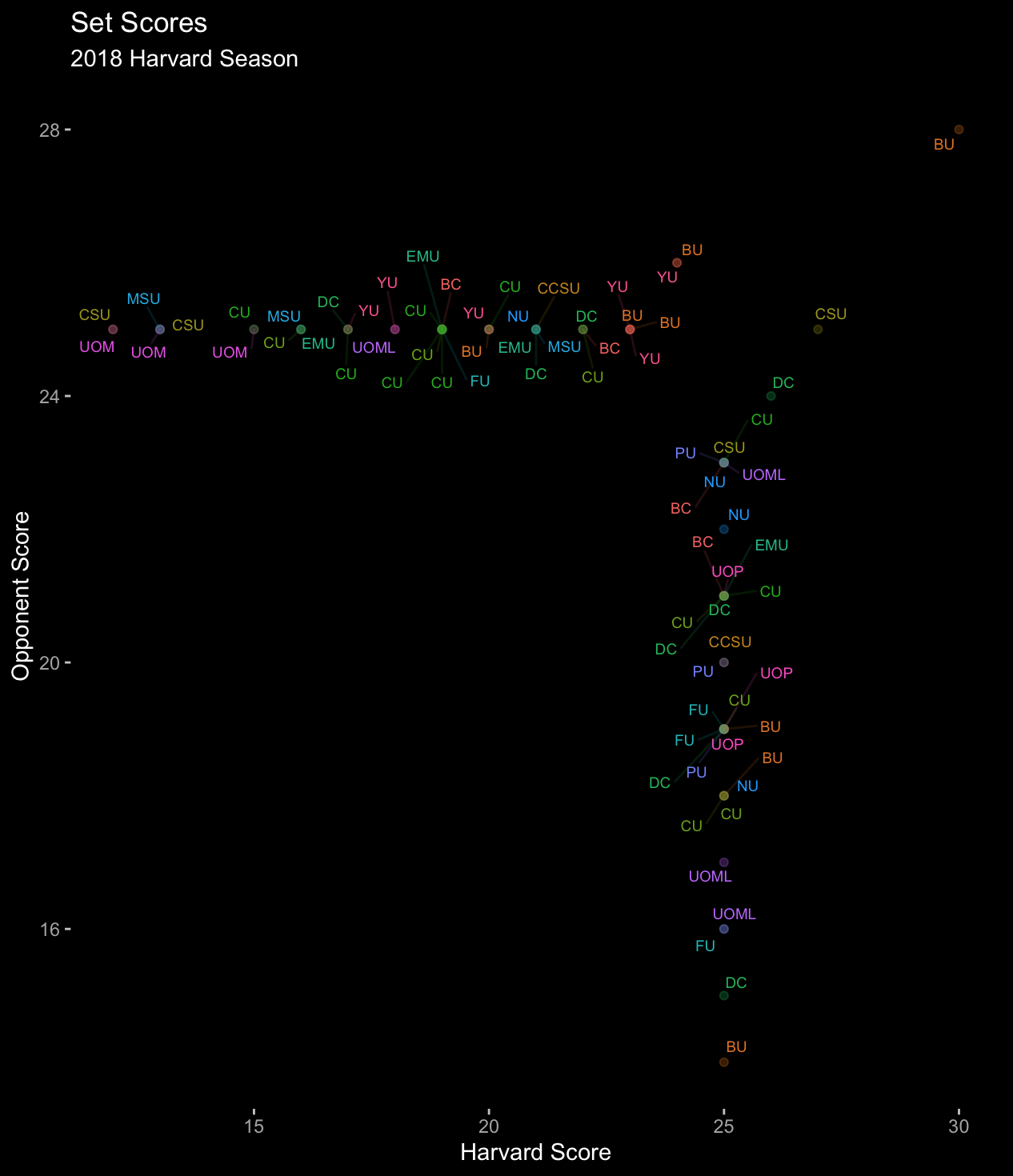
The dataset has data from 21 matches throughout the 2018 season. Each play has data regarding which player completed the action, what the action was, an evaluation of the action, start and end coordinates of the ball, and information about points, among other things. There are over 80 variables in the dataset, but some are repetitive. One limitation of this dataset is that it is only for Harvard’s team, which will mean that the insights may not be generalizable. In my R script, I have done some initial data exploration that shows tendencies of the team.



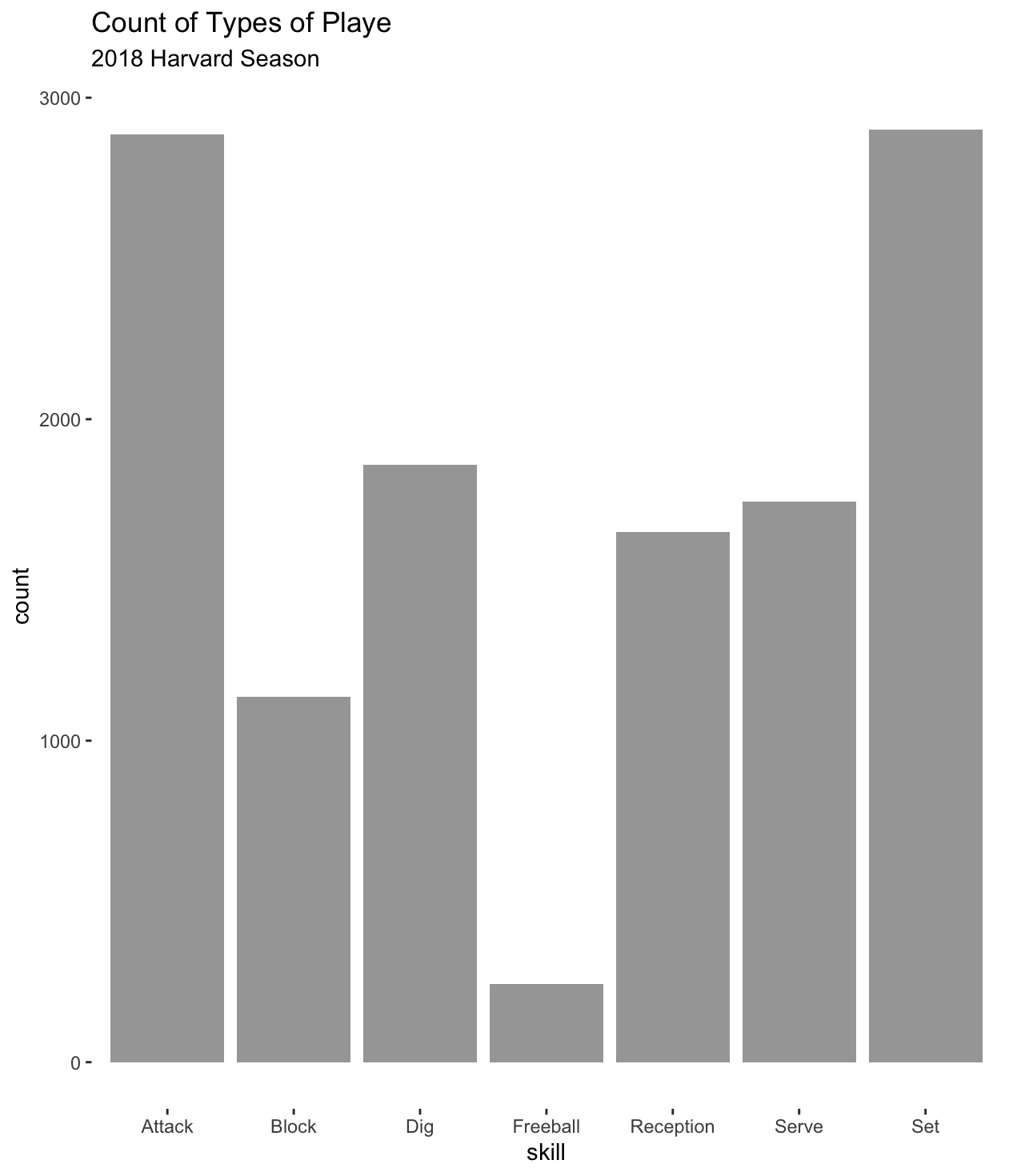
In my R script, I ran some summary statistics:





Here is one of my preliminary visualizations:

This graph shows the final score for each set (excluding games that the winning team was first to 15). Harvard’s score is on the x-axis, and their opponent’s score is on the y-axis. The points in the horizontal line near the top are sets that Harvard lost, and the points in the vertical line near the right are sets that Harvard won. Points near the top left and bottom right are sets where one team heavily dominated its opponent. From this graph, it looks like Harvard won and lost about an equal number of sets, with a larger number of sets being a close score.

Here is another preliminary visualization.

This bar chart shows a count of each type of play for Harvard’s team throughout the season. The results align pretty close to what is expected in an ordinary game. The team also does a good job at minimizing freeballs, which are usually easy for the other team to pass.