Robert Treiber III
Lacrosse Shot Chart

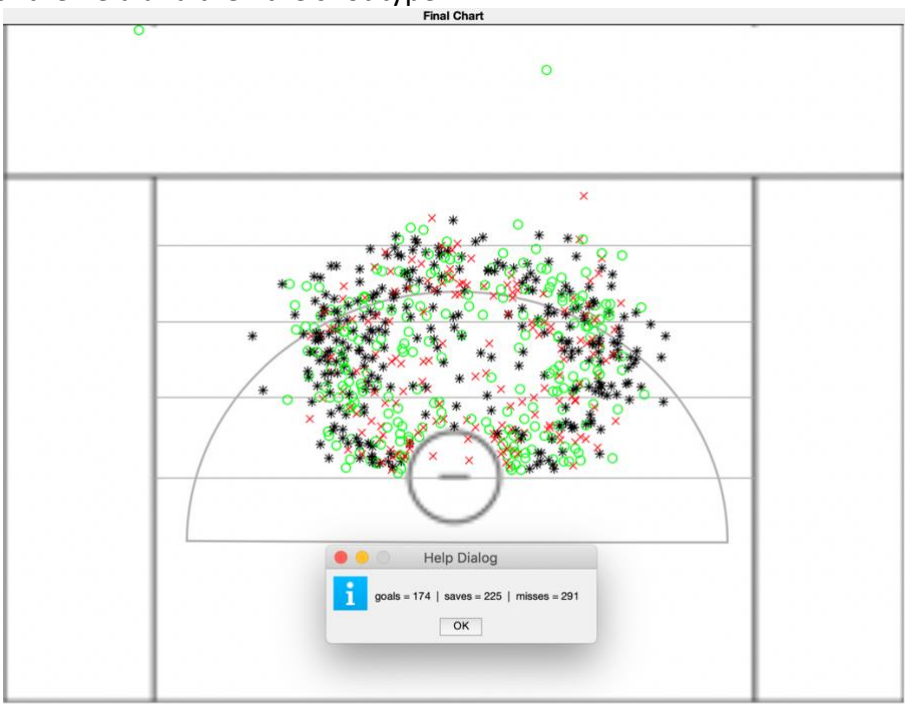
Motivation for Dataset Creation

This dataset was created over the course of an entire lacrosse season. I serve as a graduate assistant goalie coach for the Tufts Men's Lacrosse Team and during the game am responsible for tracking opponents' shot location and results. I have always been interested in baseball and basketball sabermetrics, a combination of statistics and mathematical analysis, in the form of shot charts, strike zone charts, and any metrics that can produce heatmaps. Currently, the NBA and MLB are leading the push for data driven decisions for players and coaches. Moneyball is a famous book that tells the tale of how the Oakland Athletics, a small market baseball organization in the MLB, used sabermetrics to compete with premier market teams, like the Boston Red Sox and New York Yankees.

My data was originally recorded, during the games, by hand and needed to be translated into digital data that could be cleaned, manipulated, and analyzed. Because I needed to create my own data entry program, I tried to create a program that could be manipulated to be used for any sport, not just lacrosse. Ultimately, I wanted to see if sabermetrics could be applied to lacrosse and help with defensive decision making and game planning.

Dataset Composition

One of the main differences between lacrosse and basketball is that in lacrosse when there is a shot, there is an addition possible outcome to a goal or miss, a save. After entering each game's data individually, I was able to compile the data into one shot chart for the season and calculate the total saves, goals, and misses totaling 690 different shots. Each shot has an x & y location of the field and then the shot type.



Data Collection Process

The data was collected in real-time by me during the games. A lacrosse game consists of four, fifteen-minute quarters, and overtime, if necessary. Each game I would print a new shot chart and record the location and type of each shot as it occurred. To this date, our season has included eighteen games, which includes the regular season, league playoffs and the first round of NCAA playoffs. It should be noted that blocked or deflected shots were counted as missed shots. Ultimately, the data represents all of the shots that opponents took while trying to score on us, to this date. I believe that there is some possibility for error/ noise when looking at the shot locations. While I did my best to approximate the exact location of the shot, there is a high probability that they are not recorded or entered into the program exactly where the shot occurred. In the future it, might be worth re-watching the games to confirm the shot locations because the camera that is used to film has a high vantage point than where I am positioned on the sideline.

Data Processing

During the data entry portion of my project, each game was saved individually as raw data consisting of the x & y pixel location of the shot on image and the type of shot. Cleaning the data was required to translate the pixel location of the shot to a distance in yards from the net. I used MATLab to enter/ convert my data from pen and paper into digital data and to analyze the data through machine learning toolboxes.

Dataset Maintenance

The dataset will continue to be updated as long as I am coaching. I will offer this program and data to the current and future coaches if they would like to continue to build, use and analyze the data. I will also offer my program online to others so that they may improve it and use it for their own analysis.