Player Tracking for Division I College Hockey

Maxime Bost-Brown[1], Michael E. Schuckers[2]
[1]Maxime Bost-Brown's Affiliation, St. Lawrence University, Canton, NY 13601
[2]Michael E. Schuckers's Affiliation, St. Lawrence University, Canton, NY 13601

Abstract
The purpose of this project is to analyze data from a Division I collegiate hockey team, the St. Lawrence University (SLU) Saints. Using video footage of multiple games from the 2016-17 season, students from the St. Lawrence University Sports Analytics Club recorded shot attempts by SLU women's team. For each shot several metric were recorded including shooter, outcome and (x,y) location.

Key Words: sports analytics, Hierarchical Cluster, heat map, SLU, St. Lawrence University, vizualization

1. Introduction

Within recent decades, statistics has made an increasingly significant presence in sports. The developing relationship between the two, statistics and sports, has not been the most successful. In fact, several of those who have made a career in the sports industry do not approve of statistics as being a "real" measurement of talent when it comes down to player tracking.

Player tracking consists of keeping track of several sport specific player metrics, or measures of specific information about a player (for example, the height at which a player can jump). Those who have built a career around scouting players based on the abilities that they subjectively conclude are skills worthy of recruitment do not use technology or any specific replicable documented methodology to draw conclusions about players. This is where statistics comes in. Statistics takes those recorded player metrics, and draws conclusions based only on those data.

There are flaws to both parties. The sports veteran considers (non-numerical driven) skills of a player, compared to other players, in addition to outside factors when drawing conclusions about players. Statistics draws conclusions about players considering only the numerical data recorded specific to each player, and does not consider any kind of outside factors. It is imperative to blend the two ideals together in order to achieve a superior way of categorizing a player relative to other players.

The St. Lawrence University Sports Analytics Club and the St. Lawrence University Woman's Division I Hockey Team worked together to review several game data from the St. Lawrence Hockey Team. From that game data several metrics were taken about individual players and the outcome of their attempted shots on goal. From there conclusions were drawn based solely on the data about the St. Lawrence University Woman's Hockey Team.

2. Methods

The St. Lawrence University Woman's Hockey Team provided four different game videos. Each full game video was cut into several clips of Saints shot attempts on goal. This procedure was carried out for four games. The opponents of those games were Yale, Dartmouth, Princeton, and Harvard. These four videos were accessed on hudl.[1]From there several metrics were recorded manually by the St. Lawrence University Sports Analytics Club. For each shot, the opponent name, shooter number, shot outcome, shot height, goalie obstruction, goalie side, shot x-location, shot y-location, shot generation, tracker initials, period of game, strength of opponent, and other St. Lawrence players on the ice, were recorded (by estimating a general location.

All of the analysis done on the data was conducted using the R-language on the R-studio server. In R, the specific packages used were JPEG[2], ggplot2[3], grid[4], ggmap[5], and graphics[6]. The JPEG package was used read an image in a jpeg format. The ggplot2 package (an improved version of ggplot), was used to plot statistics in a visually pleasing manner. The grid, ggmap and graphics packages were all used to create more visually pleasing graphs and heat maps.

A heat map is a "two-dimensional representation of data in which values are represented by colors".[7] In this case, spatial heat maps are utilized. A spatial heat map is a specific type of heat map. It represents two-dimensional data, paying special attention to areas where clusters, or higher densities of data points, occur.[8]

Of all of the metric taken, only the relevant columns of data were used when doing statistical analysis. Those relevant columns of data include: shooter number, result of the shot taken, and x-location and y-location of where the shoe was taken.

A cluster analysis is a method for drawing conclusions on data that are grouped together based on similarities. Similar groups of data are clustered together in their own group and dissimilar clusters are clustered together is a different group of their own.[9]

3. Application and Results

---

[1] *Hudl*, Agile Sports Technologies, Inc., 2016, www.hudl.com.

[2] Simon Urbanek (2014). jpeg: Read and write JPEG images. R package version 0.1-8. https://CRAN.R-project.org/package=jpeg

[3] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

[4] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[5] D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

[6] D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

[7] "Heat Map (Heatmap)." *Search Business Analytics*, TechTarget, July 2011, searchbusinessanalytics.techtarget.com/definition/heat-map.

[8] Dempsey, Caitlin. "What Is the Difference Between a Heat Map and a Hot Spot Map?" *GIS Lounge*, WP-PROSPERITY, www.gislounge.com/difference-heat-map-hot-spot-map/.

[9] "Data Mining -- Cluster Analysis." *Tutorials Point*, 2017, www.tutorialspoint.com/data_mining/dm_cluster_analysis.htm.

3.1 Objective

The main objective of this project is to gain more insight on Division I Woman's college hockey offensive shooting strategies against different opponents by analyzing game data. Intuitive thought would lead one to infer that the shooting strategies of a hockey team would differ when facing different opponents. This project will use the data collected from the St. Lawrence University Saints to support or disprove that inference.

3.2 Analysis

The first step in analyzing the data was to compare the distributions of percentages of the outcomes of shots taken for each individual game. Table 1 shows the percentages for shot results for each game. Figure 1 suggests that the distributions of Misses, Rebounds and Saves are similar between games against the opponents Yale and Dartmouth. The graph also suggests that the distributions of Blocks, Goals and Misses are similar across games against the opponents Princeton and Harvard. This initial plot should force one to think about the initial inference that the offensive shooting strategies of hockey teams differs when facing different opponents. The data appears to be disproving this theory. It is imperative to investigate more.

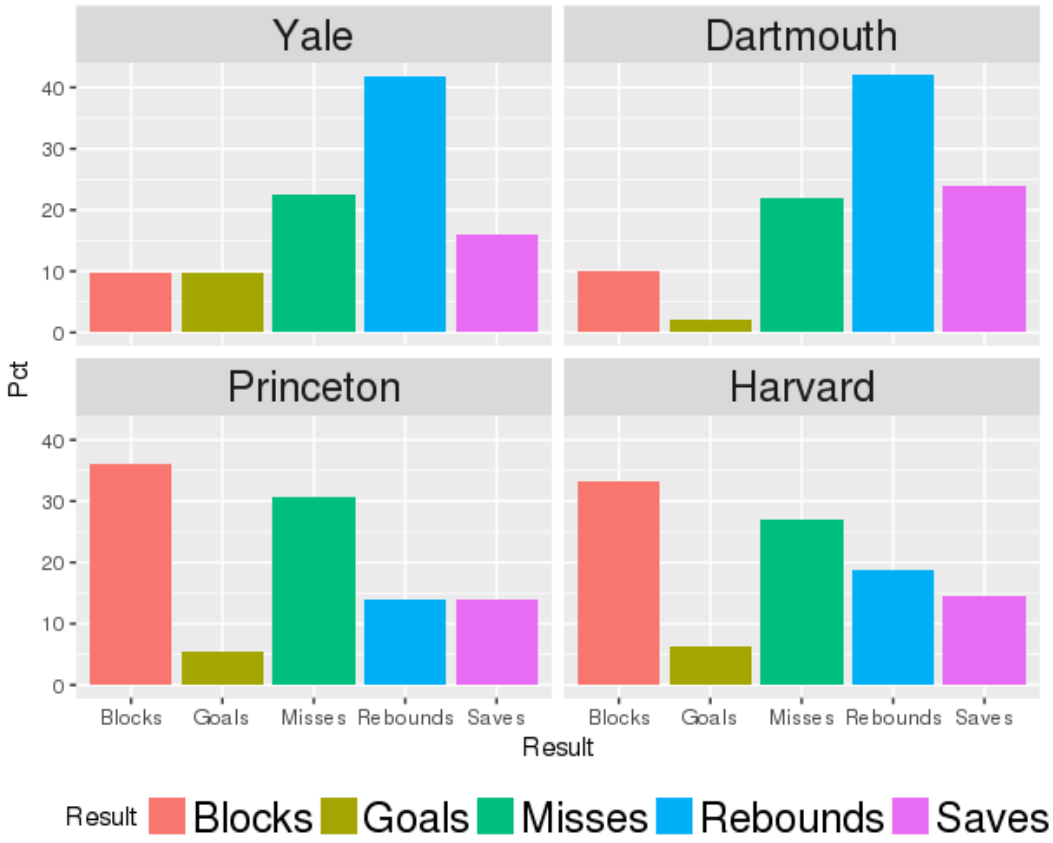

Figure 1: Plots of the distributions of shot results percentages (Blocks, Goals, Misses, Rebounds, and Saves) against opponents: Yale, Dartmouth, Princeton, and Harvard, using the ggplot2 package in R

Table 1: Shot Result Percentages Against Opponents

Opponent	Goals	Misses	Rebounds	Saves	Blocks
Yale	9.68	22.58	41.94	16.13	9.68

Dartmouth	2.00	22.00	42.00	24.00	10.00
Princeton	5.56	30.56	13.89	13.89	36.11
Harvard	6.25	27.08	18.75	14.58	33.33

For a second step, the data was plotted into a couple of different graphs. One is a type of scatter plot and another is a spatial heat map. The scatter plot of all of the shot attempts and their results does not say very much about the offensive strategies of the Saints, in fact, the plot of the data points seems random. For this reason, it is evident to use a spatial heat map to get more information.

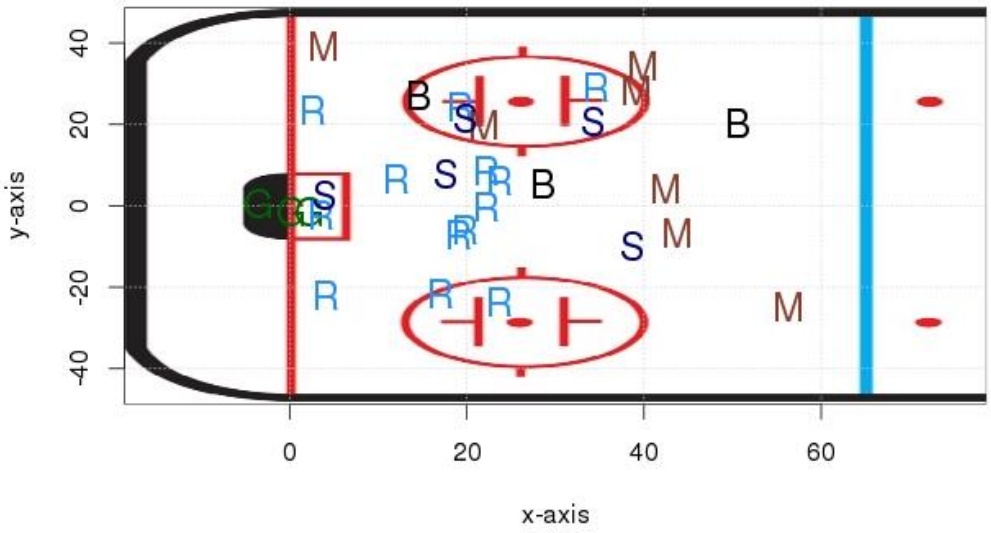

Figure 2: Yale opponent scatterplot of all recorded shot attempts on goal by the St. Lawrence University Women's Hockey Team using the ggplot2 package in R

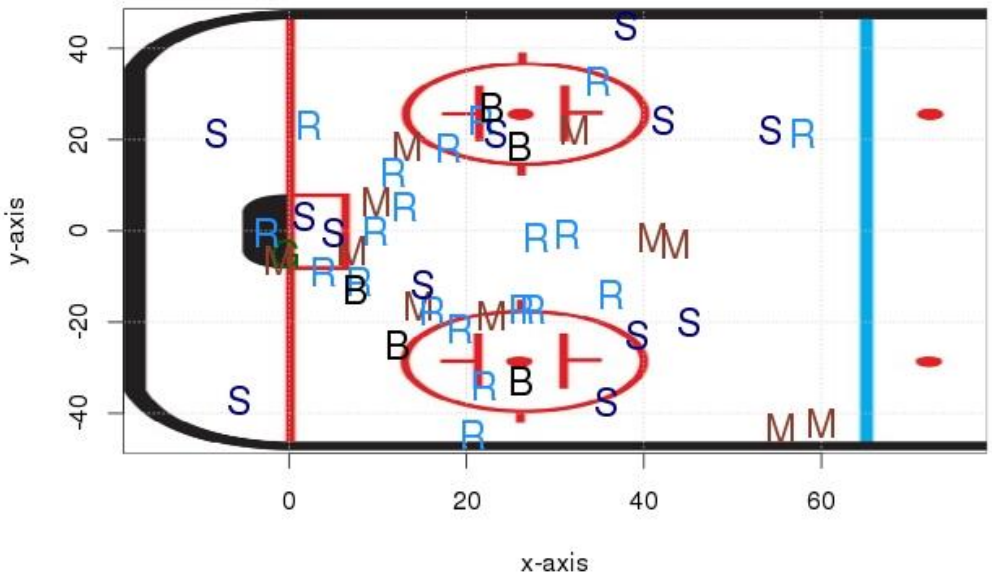

Figure 3: Dartmouth opponent scatterplot of all recorded shot attempts on goal by the St. Lawrence University Women's Hockey Team using the ggplot2 package in R

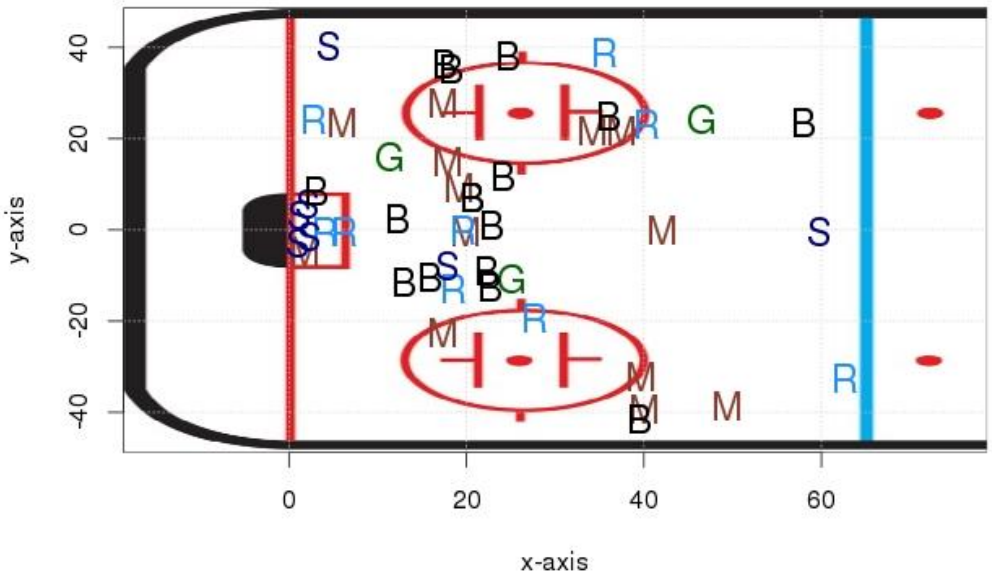

Figure 4: Princeton opponent scatterplot of all recorded shot attempts on goal by the St. Lawrence University Women's Hockey Team using the ggplot2 package in R

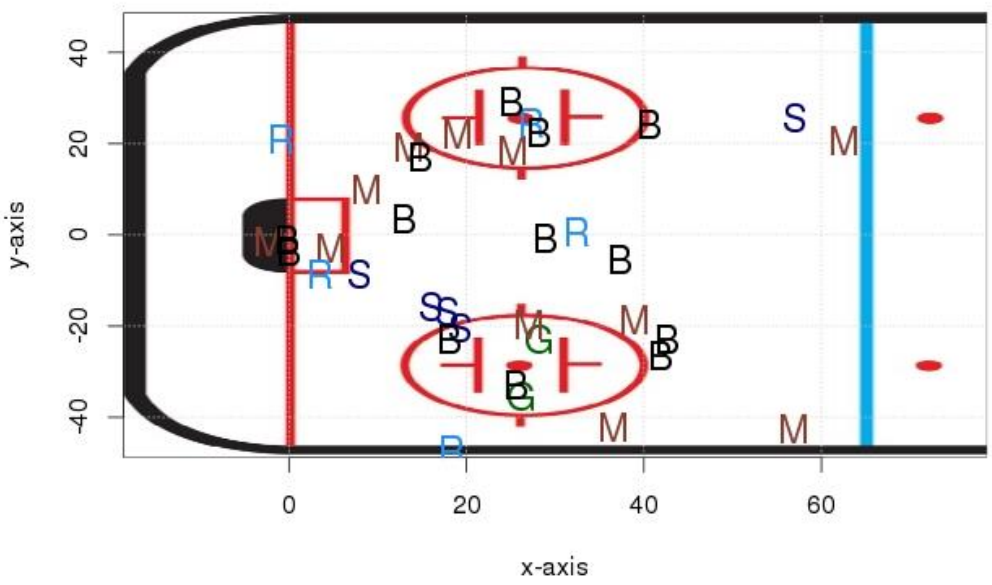

Figure 5: Harvard opponent scatterplot of all recorded shot attempts on goal by the St. Lawrence University Women's Hockey Team using the ggplot2 package in R

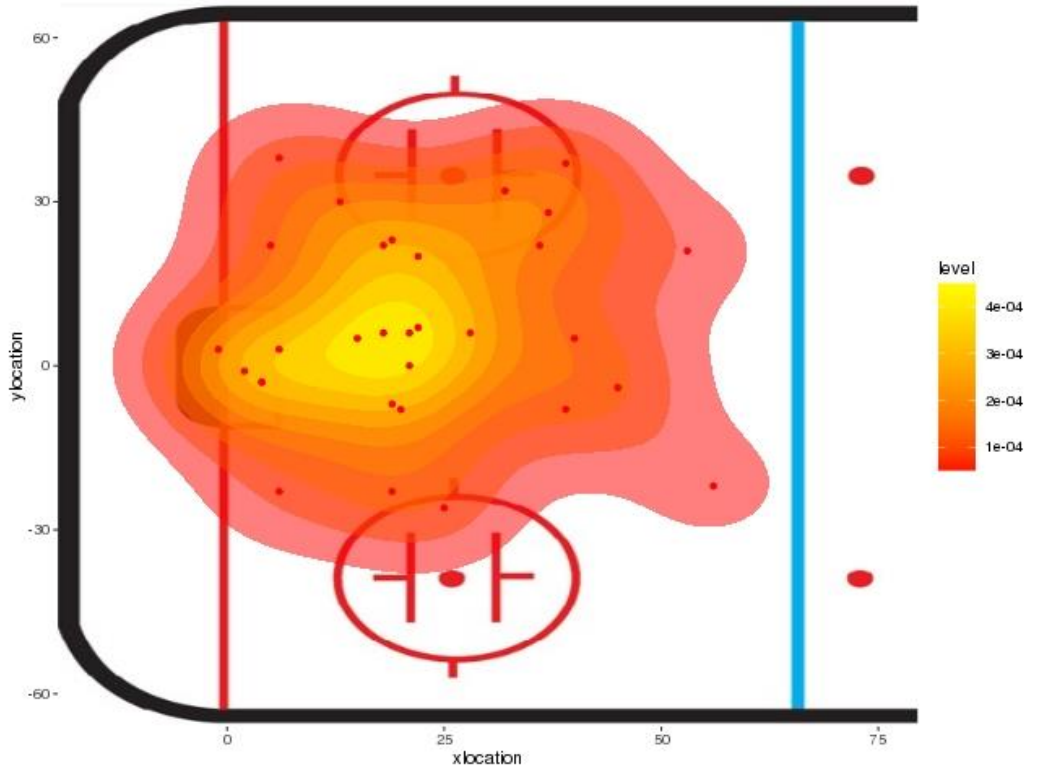

Figure 6: Yale opponent spatial heat map of all recorded shot attempts on goal by the St. Lawrence University Women's Hockey Team using the grid, ggmap, and graphics packages in R

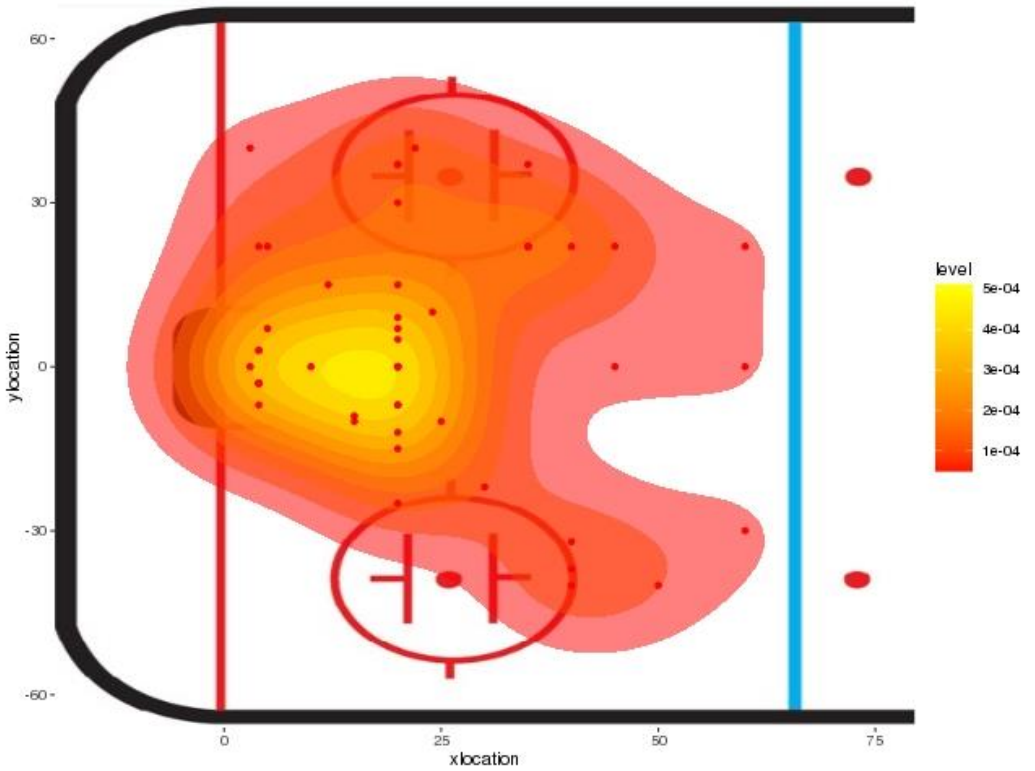

Figure 7: Dartmouth opponent spatial heat map of all recorded shot attempts on goal by the St. Lawrence University Women's Hockey Team using the grid, ggmap, and graphics packages in R

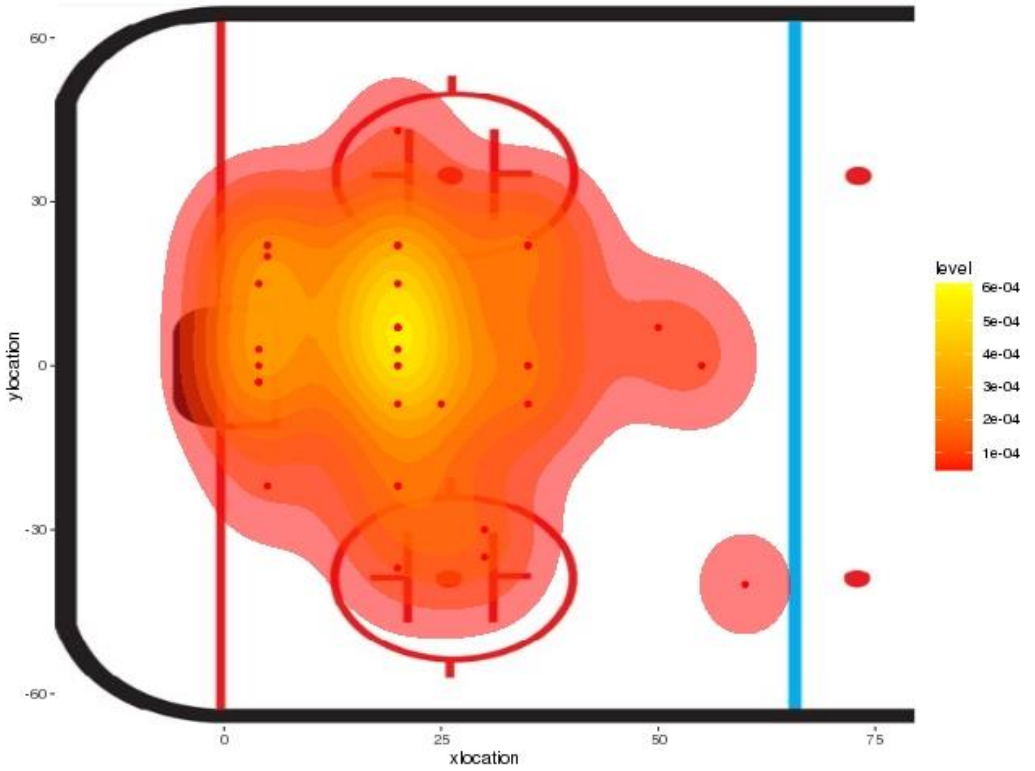

Figure 8: Princeton opponent spatial heat map of all recorded shot attempts on goal by the St. Lawrence University Women's Hockey Team using the grid, ggmap, and graphics packages in R

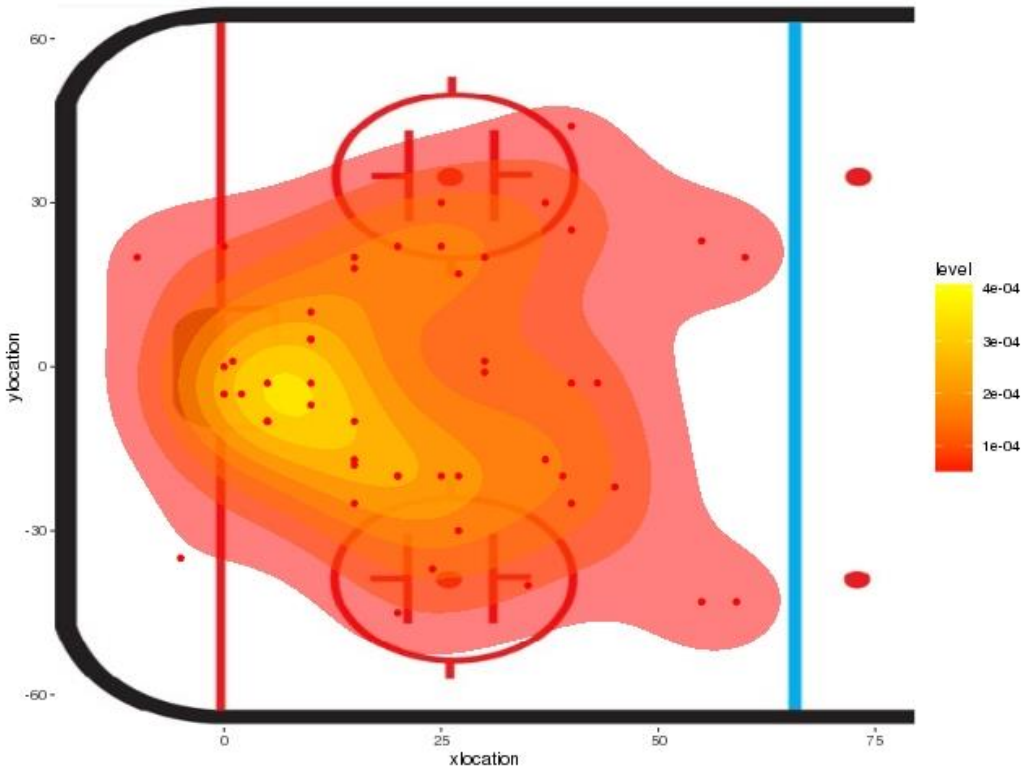

Figure 9: Harvard opponent spatial heat map of all recorded shot attempts on goal by the St. Lawrence University Women's Hockey Team using the grid, ggmap, and graphics packages in R

The spatial heat maps show interesting results that further disprove the theory that division I women's hockey teams have different offensive shooting strategies for each differing opponent. The distributions of the spatial heat maps follow closely to what the initial percentage distributions of shots indicated. The shape of the heat maps for the Yale and Dartmouth games appear to be more similar to each other than the shape of the heat maps for Princeton and Harvard games and likewise for the Princeton and Harvard games compared to the Yale and Dartmouth games. This cannot be confirmed definitively simply by observing the shapes of heat maps, so this is where the cluster analysis comes in handy.

3.3 Cluster Analysis

In this case, cluster analysis will be used to determine if the distributions of shot attempts on goal for each game are similar to one another or dissimilar. The dendrogram (refer to Figure 10 A-10) shows that the Princeton and Harvard games are clustered together and the Yale and Dartmouth games are clustered together. This numerically confirms the previous suspicions about the similarities between the shot attempt distributions of Princeton and Harvard and also Yale and Dartmouth. Since the Princeton and Harvard games are clustered together in one cluster and Yale and Dartmouth are clustered together in a different cluster, the shot attempt percentage distributions of Princeton and

Harvard are different from the shot attempt percentage distributions of Yale and Dartmouth.

4. Discussion and Conclusion

The results from this project do provide more insight on division I woman's college hockey offensive shooting strategies against different opponents. In fact, this project does not support the idea that the shooting strategies of division I women's college hockey teams are unique to each opponent. On the contrary, it has been found that offensive shooting strategies between different opponents actually may not be different.

One issue that presented itself fairly early in this project was the amount of error within the measurements. Since many of the player metrics that were recorded were on video, the only way to record the data was to estimate, for example, the general location of where a shot was taken. This relative estimation is where the error lies. The use of modern technology (like to matric tracking hockey pucks the NHL have introduced) can easily combat this concern.

As mentioned in the introduction, it is important to blend together the two ideas (subjective theories and data driven theories) in order to create a superior way of evaluating players and their teams. In the section above, the results are all data driven. There are no subjective points of views or inferences made. There is only data, and the results that the data display. The biggest result that rings here is that the shooting strategies of a team does not necessarily differ from opponent to opponent. Now, this result does not indicate as to why this is the case or if there are any underlying factors or confounding variables or anything. This is where this project lacks (and any other sports analytics project that does not involve subjective inquiries). Perhaps if there was another perspective from a source who built a career out of subjectively studying players, there would be more of a way to apply an answer to the, "why" behind the results.

Only working together, both with data and with experience, will provide the boost that the sports analytics industry needs in order exponentially soar and face new challenges.

Acknowledgements

This research would not have been possible without the opportunities presented by St. Lawrence University. I am grateful for the continued support presented to me and wish to one day return the favor to a future Laurentian.

I would also like to show my gratitude to Dr. Michael E. Schuckers and the Math, Computer Science, and Statistics Department of St. Lawrence University for introducing me to the world of sports statistics and setting me down a path to watch as the filed of statistics changes the world.

References

Hudl, Agile Sports Technologies, Inc., 2016, www.hudl.com.

"Heat Map (Heatmap)." *Search Business Analytics*, TechTarget, July 2011, searchbusinessanalytics.techtarget.com/definition/heat-map.

Dempsey, Caitlin. "What Is the Difference Between a Heat Map and a Hot Spot Map?" *GIS Lounge*, WP-PROSPERITY, www.gislounge.com/difference-heat-map-hot-spot-map/.

"Data Mining -- Cluster Analysis." *Tutorials Point*, 2017, www.tutorialspoint.com/data_mining/dm_cluster_analysis.htm.

Simon Urbanek (2014). jpeg: Read and write JPEG images. R package version 0.1-8. https://CRAN.R-project.org/package=jpeg

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf