Getting Ready for Puck Drop: Explaining and Predicting Hockey through Data Visualizations

Nataly Beribisky

PSYC6135

Introduction

Ice hockey is a fluid and thrillingly fast-paced game. Winning a hockey match, depends not only upon the contribution of individual player's strengths, but relies also on the effects of team and line mates, level of competition, load management, and seemingly many other innumerable factors. For this reason, at first glance, hockey and its outcomes appear to be exceedingly unpredictable. Indeed, if there were formulaic strategies to winning games, how could a team like the Toronto Maple Leafs, with seemingly endless funds and modifications to the team's roster and management, have not won the National Hockey League (NHL) playoff championship's Stanley Cup since 1967 (National Hockey League, n.d.)?

However, consider a hockey pass called, "the Royal Road". The concept of the Royal Road was introduced by Steve Valiquette, a former goaltender in the NHL (NHL Numbers, 2018a). As shown in Figure 1, the Royal Road imaginarily reaches between the middle of the goalie net and the midpoint of the top of the faceoff circles (NHL Numbers, 2018a). A Royal Road pass is argued to be the most dangerous pass players can make in the attacking zone since it requires the opposing team's goaltender to move sideways, creating an opening in the net where there was not one before. How dangerous is the Royal Road Pass? The Passing Project, an initiative where volunteers that tracked all the passes in the NHL that forced a shot, found that a Royal Road pass generated shots on goal 27.80% of the time (NHL Numbers, 2018a).

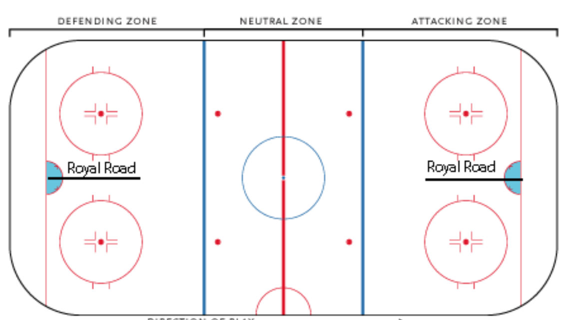

Figure 1. Ice rink with marked royal road pass. Retrieved from https://www.nhlnumbers.com/2016/08/10/passing-project-dangerous-primary-shot-contributions

Concepts such as the Royal Road are noteworthy in that they make the game of hockey more quantifiable and arguably more exciting. If teams become aware that the Royal Road pass is the most threatening pass in the offensive zone, then a team in the offensive zone should try to increase instances of Royal Road passes. In contrast, the team in the defensive zone should try to avoid instances of Royal Roads as much as possible, by blocking the area with defensive players. The incorporation of the Royal Road example here is to show the potential power of quantifying the game of hockey and using statistics to understand it. On a broader level, in 1998, Williams and Williams stated that statistical applications to hockey appeared to be scarce but growing, likely in tandem with the popularity of the sport. More recently, Swartz (2017), stated that most NHL teams now employ analytics teams, some even headed by academic statisticians.

Accordingly, hockey, with its beginnings traced back to 19th century cold Canadian winters, frozen ponds and chaotic ever-changing rules (Williams & Williams, 1998), presently appears just as enjoyable, yet more structured and consistent, making the game more amenable to data analysis. A key aspect of analytics and understanding and predicting the hockey is data visualization. Visualizations have a role in simplifying and communicating multi-faceted information. Visualizations are valuable for hockey specifically in that they can show both changes in the game over time while also presenting the multivariate outcomes of hockey (e.g., shots, goals, assists, goals against). Most importantly, data visualizations can help present hockey easily, and make the sport accessible to viewers and teams in a different way than simply watching the game. This paper aims to underline the helpful role of data visualizations in both explaining and understanding hockey and making predictions about it. At the very least, the reader may have gained insight into Royal Road passes, which hopefully makes their hockey watching experiences more entertaining.

Hockey Visualizations to Explain and Understand the Game

Visualizations presented in this section of the paper allow for a deeper understanding of hockey. They range from shot tide visualizations, which give a temporal summary of the game, to 5v5 match ups, which show which team players play together against specific competition, to the use of different passing methods. These visualizations help understand the concepts of matching competition, game flow, and different passing techniques.

Shot Tides

Figure 2 is a shot tide visualization by Micah Blake McCurdy (2019a). Shot tides are an elegant way of presenting each team's shot pressure in the offensive zone. The visualization is separated by a vertical line to create two shot tides, one for each team. The tides are also split into three pieces by two horizontal lines, denoting the three 20 minute periods. Shot tides are read from top to bottom: specifically, time starts at the team labels, and ends at the "Shots per 60m" label.

Each label in the shot tide represents the name of a player who scored a goal. In Figure 2, the Toronto Maple Leafs won, scoring 3 goals over The Boston Bruins' 2. The darker blue areas demonstrate when the team was

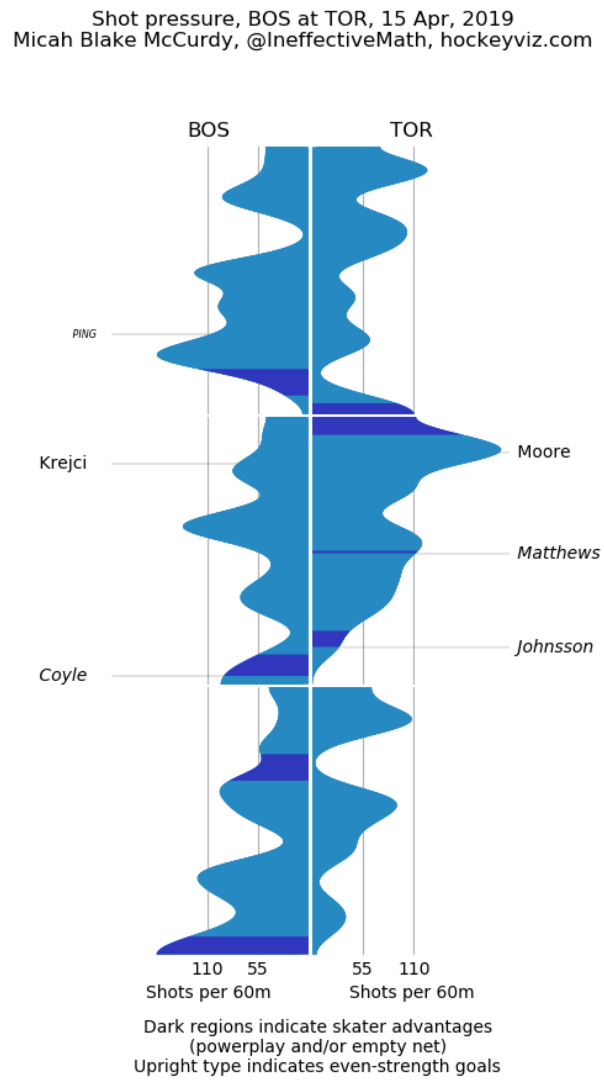

Figure 2. Example of a shot tide visualization. Retrieved from https://hockeyviz.com/game/2018030123/tide

playing with an additional advantage (e.g., a power play or pulling the goalie). In this instance, at the bottom of the shot tide, it is possible to deduce that the Bruins pulled their goalie (for another offensive player), to increase their chances of scoring an additional goal to tie the game. If many shots are created in a short time-frame the area of the shot tide is large. For example, in Figure 2, the Bruins generated more shots than the Maple Leafs in the last half of the third period because they were a goal behind.

Another potential incorporation into shot tide visualizations may be to label each minute of the game on a Y-axis to show the exact minute a goal was scored. However, even without this additional information, shot tides present shot chances, team advantages, and game outcomes in a simple and well-designed way.

5v5 Matchups

5v5 matchups, also from McCurdy (2019b), serve as a way to visualize which players play together during a hockey match. To explain the use and structure of this visualization, it is useful to briefly introduce some hockey terminology and rules.

Offensive players play in groups of three called “lines,” while defensive players have a partner and play in “pairs.” As there are varying skill levels within team, the first (or best) line of a team is usually matched to the first line and strongest pair of the opposing team to avoid overly-increased scoring chances by the top line.

The 5v5 visualization can be used to investigate the line matching in a specific hockey game. Figure 3 shows an example of the 5v5 visualization of the same game where the Maple Leafs won 3 to 2, over the Boston Bruins. The visualization is split into something like a matrix, with individual Toronto players as row names, and individual Boston players as column names. Each entry in the matrix uses the square’s (a) size, (b) colour proportion, and (c) brightness to represent, (a) the proportion of ice time the two players shared, (b) the proportion of shots made by each player, and (c) the number of shots those two players made on the ice (or the amount of positive activity generated in either offensive zone).

This visualization allows the viewer to deduce player line ups and match ups quite quickly. For instance, squares of certain size and brightness always seem to co-occur together. In Figure 3, for Toronto, one can see that not only are A. Johnsson, A. Matthews, and K. Kapanen, line mates, but for this game, they were constantly matched against the Boston line of J. DeBrusk, D. Krejci, and K. Kuhlman. Further, all nine squares here are equally bright, indicating that both lines made many shots. Indeed, looking back to Figure 2, three out of the five goals scored in the entire game were made by players from this match up. It is also possible to deduce that this Toronto line was consistently playing against a specific defensive pair, T. Krug and B. Carlo, as this Toronto line has the largest squares with this Boston pair than any other pair.

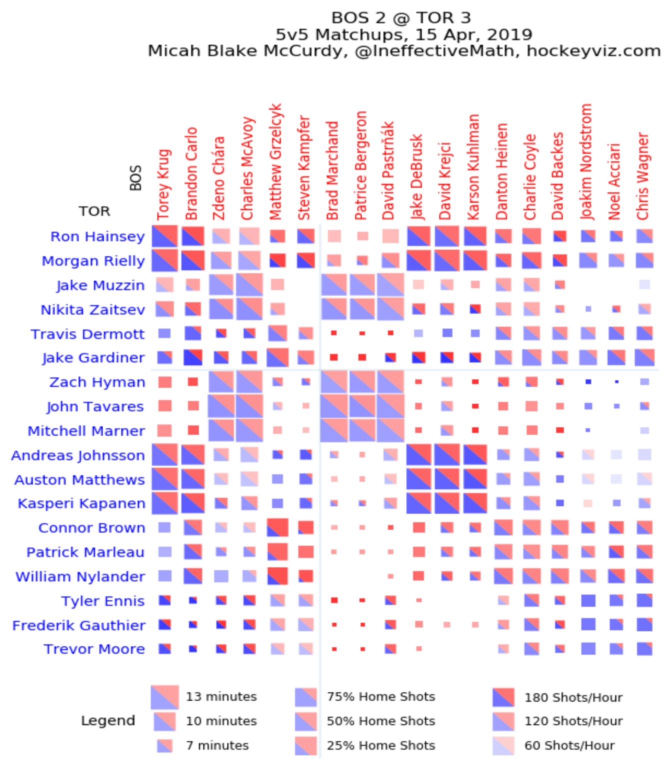

Figure 3. Example of a 5v5 matchup visualization. Retrieved from https://hockeyviz.com/game/2018030123/matchups

One of the strongest elements of this visualization is its organization. Specifically, it would not be possible to see the line matches and team matches as clearly, if the author had not organized the matrix by each team's offensive lines and defensive pairs. This visualization is a useful tool for both spectators of the sport to deduce line matches, or even for teams to make decisions about pair and line matching.

The Passing Project Chart

The former two visualizations presented were both ways of describing any specific game. Figure 4, a visualization which is part of the Passing Project, allows a comparison of a specific tactic to other comparable ones during a hockey game (NHL Numbers, 2018a).

Each column heading in Figure 4 is a type of pass made in the offensive zone: behind the net, faceoff (i.e., in one of the offensive circles), other, point (i.e., from the back of the offensive zone), and the aforementioned Royal Road pass. Each circle represents the following shot after the pass: orange ones are scoring chances, and blue ones are not. This visualization gives compelling evidence that behind the net and Royal Road passes generate more scoring chances than other types of passes. In contrast, passes from the point are least effective for creating scoring opportunities.

This visualization is an effective use of contrasting colours and presenting considerably large data in a clear way to

Figure 4. Different offensive zone passes and scoring chances. Retrieved from https://www.nhlnumbers.com/2016/08/10/passing-project-dangerous-primary-shot-contributions

convince the viewer. Unfortunately, it was not possible to find an accompanying supporting definition of a scoring chance. Although some defining it at as an attempted shot at goal made from the top of the face-off circles, there appears to be no conclusive definition (NHL Numbers, 2018b). This visualization would therefore be strengthened by including an operational definition of a scoring chance, and perhaps splitting scoring chances into high danger (high probability of goal incidences) and low danger opportunities (low probability of goal incidences).

Hockey Visualizations to Predict the Game

The two visualizations presented in this section are used to predict game outcomes. Predictions may be used both in the regular season, to estimate how many games a team may win, and to predict playoff performance to determine which team may win a playoff championship. An example of each is presented below.

Projected Points

Figure 5 is a visualization that uses boxplots to represent predicted regular season points for all NHL teams in the 2018-2019 regular season. The visualization is based on an output of total points for 50,000 simulations based on Perry's (2017) prediction model called Salad. The Salad model is a combination of sub-models that include naïve Bayesian models, logistic regressions, random

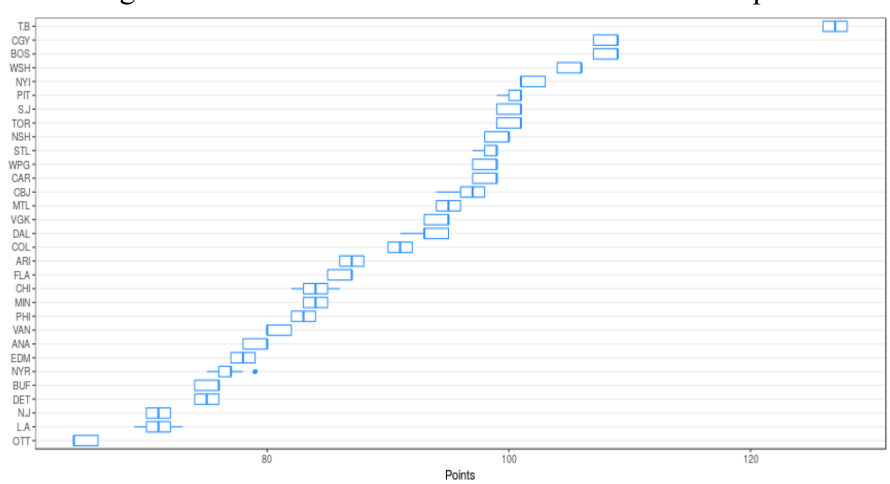

Figure 5. Regular season predictions for the National Hockey League, with each team's total points represented by a boxplot. Retrieved from https://www.corsica.hockey/predictions

forests, and neural networks (Perry, 2017). The model was created to have minimal logarithmic loss, or minimal overconfidence in incorrect outcomes. It was also cross-validated and tested on prior NHL season data.

The display of the model's predictions serves as a way to view a team's projected performance, estimating how they will finish the regular season (Corsica, n.d.). For fans, it may also serve as a way to predict whether their favourite team will make the playoff series and which opposing team their favourites may be playing during playoffs (since this matching is partially determined by points).

While Figure 5 is a useful summary of the model's data, it would be beneficial for the visualization to have included when the model was run. Although there is accompanying text that states that simulations were based off, "of the games remaining in the regular season," it was not clear at what point the model results were simulated (Corsica, n.d.). For example, Figure 5 demonstrates that the Tampa Bay Lightning were well ahead of all the other teams, projected to have a 127 point season. After the end of the regular season, it is possible to state that this estimate was quite accurate since Tampa finished with 128 points (National Hockey League, 2019). However, the viewer may judge the strength of the model and the accompanying visualization differently, if they knew whether it was run mid-way through the season, or just before the last ten games. Finally, the visualization may have been strengthened by using different colour boxes for the four divisions each team belongs to, as play-off team matching takes division into account.

Playoff Simulation Brackets

Figure 6 is an example of a playoff simulation bracket (Electronic Arts Games, 2019). Although the simulation techniques and sampling distributions used by Entertainment Arts are unknown, the visualization exposes the viewer to Electronic Arts' predictions for all rounds of the playoffs. Each column represents a different round of the playoffs, with earlier rounds closest to the edge of the visualization, and later rounds closer to the center. The visualization puts the championship final, the last round of the playoffs, right at the center of the graphic, so that the viewer's eye is immediately drawn to the most important game and its projected winner (shown by the score which represents the number of games won). Other than containing the number of games each team is projected to win, the circular logos within each square represent each playoff team. The box around each matchup is the colour of the predicted winning team's logo.

This visualization is an elegant way of viewing simulated data, while also presenting copious information about playoff structure within the NHL.

Figure 6. An example of a playoff simulation bracket. Retrieved from https://www.ea.com/games/nhl/news/nhl-19-stanley-cup-playoffs-simulation

Visualizing the Game

The aim of this paper has been to show the multitude of ways that hockey can be visualized. From line matching, to predicting playoff winners, visualizations allow for large amounts of data of many types to be presented accessibly. Ideally, visualizations within hockey organizations serve to help teams make critical decisions about the success of specific tactics and matchups. For fans, visualizations can be a useful way to engage with hockey data and learn about the sport, outside of watching the game. The knowledge gained from viewing game data in these ways can be applied to watching hockey, which in turn, makes looking at next day visualizations of games even more exciting.

References

Corsica. (n.d.). Predictions. Retrieved from http://corsica.hockey/predictions/

Electronic Arts (2019). Playoff simulation. Retrieved from https://www.ea.com/games/nhl/news/nhl-19-stanley-cup-playoffs-simulation

McCurdy M. B. (2019a). Shot pressure: BOS @ TOR, 15 Apr 2019. Retrieved from https://hockeyviz.com/game/2018030123/tide

McCurdy M. B. (2019b). 5v5 matchups: BOS @ TOR, 15 Apr 2019. Retrieved from https://hockeyviz.com/game/2018030123/matchups

National Hockey League. (n.d.). All-time NHL results. Retrieved from https://www.nhl.com/info/all-time-nhl-playoff-results

National Hockey League. (2019). Statistics. Retrieved from http://www.nhl.com/stats/team?reportType=season&seasonFrom=20182019&seasonTo=20182019&gameType=2&filter=gamesPlayed,gte,1&sort=points,wins

NHL Numbers (2018a). Passing project: Dangerous primary shot contributions. Retrieved from https://www.nhlnumbers.com/2016/08/10/passing-project-dangerous-primary-shot-contributions

NHL Numbers (2018b). Why the “Scoring Chance” is a flawed metric, and how we can improve on it. Retrieved from https://www.nhlnumbers.com/2016/11/16/why-the-scoring-chance-is-a-flawed-metric-and-how-we-can-improve-on-it

Perry, E. (2017). On salad and predicting hockey games. Retrieved from http://www.corsica.hockey/blog/2017/10/06/on-salad-and-predicting-hockey-games/

Swartz, T. B. (2017). Hockey analytics. *Wiley StatsRef: Statistics Reference Online, To appear.*

Williams, W. B. H., & Williams, D. A. (1998). Performance Indices for Multivariate Ice Hockey Statistics In J. Bennett (Ed.), *Statistics in sports* (pp. 141 -155). New York, NY: CUNY.