Introduction

Royal road passes (RRPs) have been a commonly discussed concept in hockey tactical analysis over the past five years (Valiquette, 2015). These passes are known to be very effective in leading to goal scoring opportunities, which make them highly sought after for offenses, but also heavily guarded against by defenses. Although common knowledge suggests the value of these passes is derived from factors like the goaltender needing to move from post to post, or the closeness of the resulting shot, the specific variables that uniquely contribute to the value of RRPs remain to be clarified. Elucidating these relationships could help offenses find alternative routes for scoring when the opposing team tightly defends the royal road. The present study examines factors that make RRPs valuable and explores other passing patterns that may lead to similar or better outcomes. Because RRPs often force goaltenders to move far distances to save, I hypothesized that pass-to-shot angle change would be among the most influential variables in predicting scoring, along with shot distance, which is often a highly weighted variable in published expected goal (xG) models.

Method

The dataset for this analysis was obtained from Stathetes through the Big Data Cup 2021 and contains event level data from 53 games from the OHL, NCAA, and Olympics. Data analysis was conducted using R version 4.0.3. The data was transformed to map the pass location to the subsequent shot. Because quickly taken shots presumably have a higher chance of scoring, only one-timers were included in the analyses as the dataset did not include the amount of time between receiving a pass and shooting, using only one-timers allowed for the isolation the spatial effects of RRPs. Given that the dataset contained 880 different instances of one-timers, it was impractical to visualize and categorize them all by hand. Thus, a custom function was built to identify line-line intersections between the pass, and imaginary line segments that represent the royal road, stretching from (142, 39.5) to (189, 39.5) and from (142, 45.5) to (189, 45.5). The passes that crossed both of these horizontal lines were categorized as RRPs.

Variable Selection

To measure the effectiveness of each pass/shot combination, the coordinates of the pass and shot were used to create four variables. (1) Shot Distance represent the distance, in feet, between where the shot took place and the center of the net. (2) Pass Distance represent the distance, in feet, between where the pass and shot took place. (3) Shot Angle represent the angle closest to the shooter in a triangle formed between the shooter and both posts, measured in degrees. The larger the angle, the larger the shooting area would appear to the shooter, goaltender notwithstanding. (4) Change in Angle from Pass to Shot represent the angle where the center of the goal is one point of a triangle formed between the passer and shooter, measured in degrees. This angle is a proxy for how far the goaltender has to move laterally in order to make a save, with larger angle indicating larger movements necessary for the goaltender to save. These four variables were created to isolate the variables relevant to RRPs, and the angle measurements adjust for potential biases that may exist in the dataset due to shooter handedness. Figure 1 represents histograms of the four variables, separated by RRP categorization.

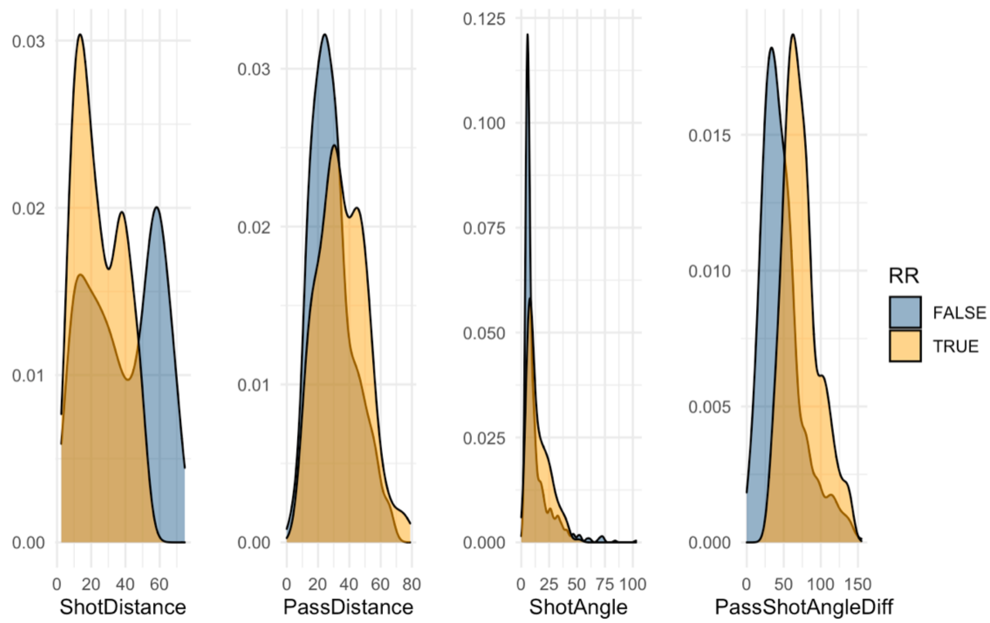

Figure 1: Histogram of variables across RR categorizations

<u>Expected Goals (xG) Model Building</u>

Next, these variables were run through several different classification models in order to find the best fitting xG model. In accordance with previous work in the field (e.g., MacDonald, 2012; Younggren & Younggren, 2018; Novet, 2019), the strength of the model was evaluated by measuring area under the curve (AUC) of the receiver operating characteristic curve (ROC). This approach standardizes model output that allows for comparison against other similar models available.

<u>Relative Variable Importance</u>

Within the resulting xG model, variable importance was examined using the vip package in R, with higher value indicating a higher percentage of variance explained in the model by a given variable.

<u>Alternative Passing Patterns</u>

Finally, to explore other passing patterns that create similar or better scoring outcomes, the top 10% of non-RRP shots were selected for visual examination of passing patterns. This subset of non-RRP shots was then compared with RRP shots using an independent-samples t-test on xG, with and without relevant covariates.

Findings

The Expected Goals (xG) Model

The best-fitting xG model was a boosted random forest model, using the engine xgboost. This approach is almost identical to what was used by Josh & Luke Younggren in their xG model write up (Younggren & Younggren, 2018). While the Younggrens defined their hyperparameters by hand through a modified grid search, the current project automatically selected the hyperparameters through the tune package in tidymodels. Although this approach makes the end result harder to interpret, this xG model results in a tool to evaluate passes, rather than a player and team evaluation tool on a highly trafficked website, like Evolving-Hockey.com. The hyperparameters chosen resulted in a ROC AUC of .774 (fig. 2), which was better than expected, and far stronger than the simple logistic regression models that started this process (ranging from .25 - .35 ROC AUC).

```
══ Workflow ═══════════════════════════════════════
Preprocessor: Formula
Model: boost_tree()

── Preprocessor ───────────────────────────────────
IsGoal ~ .

── Model ──────────────────────────────────────────
Boosted Tree Model Specification (classification)

Main Arguments:
  mtry = 4
  trees = 1000
  min_n = 17
  tree_depth = 14
  learn_rate = 0.0264800933378833
  loss_reduction = 0.000452959840820337
  sample_size = 0.452340360549279

Computational engine: xgboost
```

Figure 2: Hyperparameters chosen by tune function

Relative Variable Importance

Contrary to hypotheses, analysis of the relative importance of the variables showed that Shot Distance (0.38) and Shot Angle (0.36) were the most influential variables in the model (fig. 3). In other words, smaller shot distances and larger shot angles accounted for 38% and 36% of the variance in the xG model, respectively. These two variables were followed by Pass Distance (0.18). Pass Shot Angle Difference (0.06) and RRP categorization (< 0.01) had very little effect on shot outcomes.

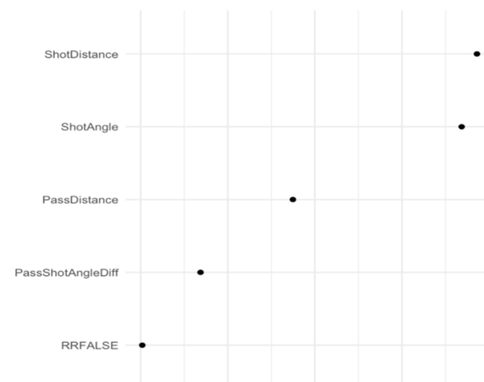

Figure 3: Relative variable importance

Alternative high xG passing patterns

To identify alternative, high xG passing patterns, non-RRP pass/shot combinations in the top 10% of xG were selected for additional analyses. Within this subset, two distinct passing patterns emerged (fig 4). The first (orange) consisted of passes that only crossed one of the lines of the imaginary rectangle that placed the shooter directly in front of the goal. The second (blue) pattern consists of passes from below the goal line, short passes to close range shots mostly centered at the top of the crease, with some coming in from the sides of the crease.

Finally, an independent-samples *t*-test compared these top-performing non-RRPs against RRPs on xG, and showed that the average xG was higher in the non-RRPs (0.17) than the RRPs (0.10), $t(179) = 21.15$, $p < .001$. This pattern remained consistent even after adjusting for Shot Distance and Shot Angle (the two most influential variables in the xG model, as described above), $F(1,245) = 61.63$, $p < .001$.

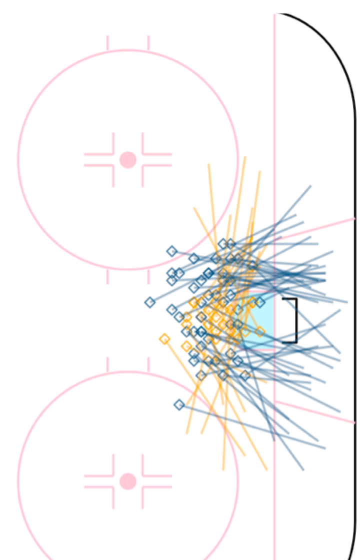

Figure 4: Top non-RRP passing patterns. Orange: "semi" RRP chances. Blue: Passes coming from below the goal-line

Discussion

The present project found shot distance and shot angle to account for the most variance in predicting goal scoring, among variables that commonly make up RRPs. This is apparent when showing the shot attempts plotted (fig 5). The relative importance of shot distance has been highlighted by Alex Novet, who noted “the single most important feature by a wide margin is how close the shot is to the net” (Novet, 2019). However, the gulf between the shot-related variables and pass-related variables was notably large, with the former collectively accounting for 74% of the variance in the model, while the latter only accounting for 24%.

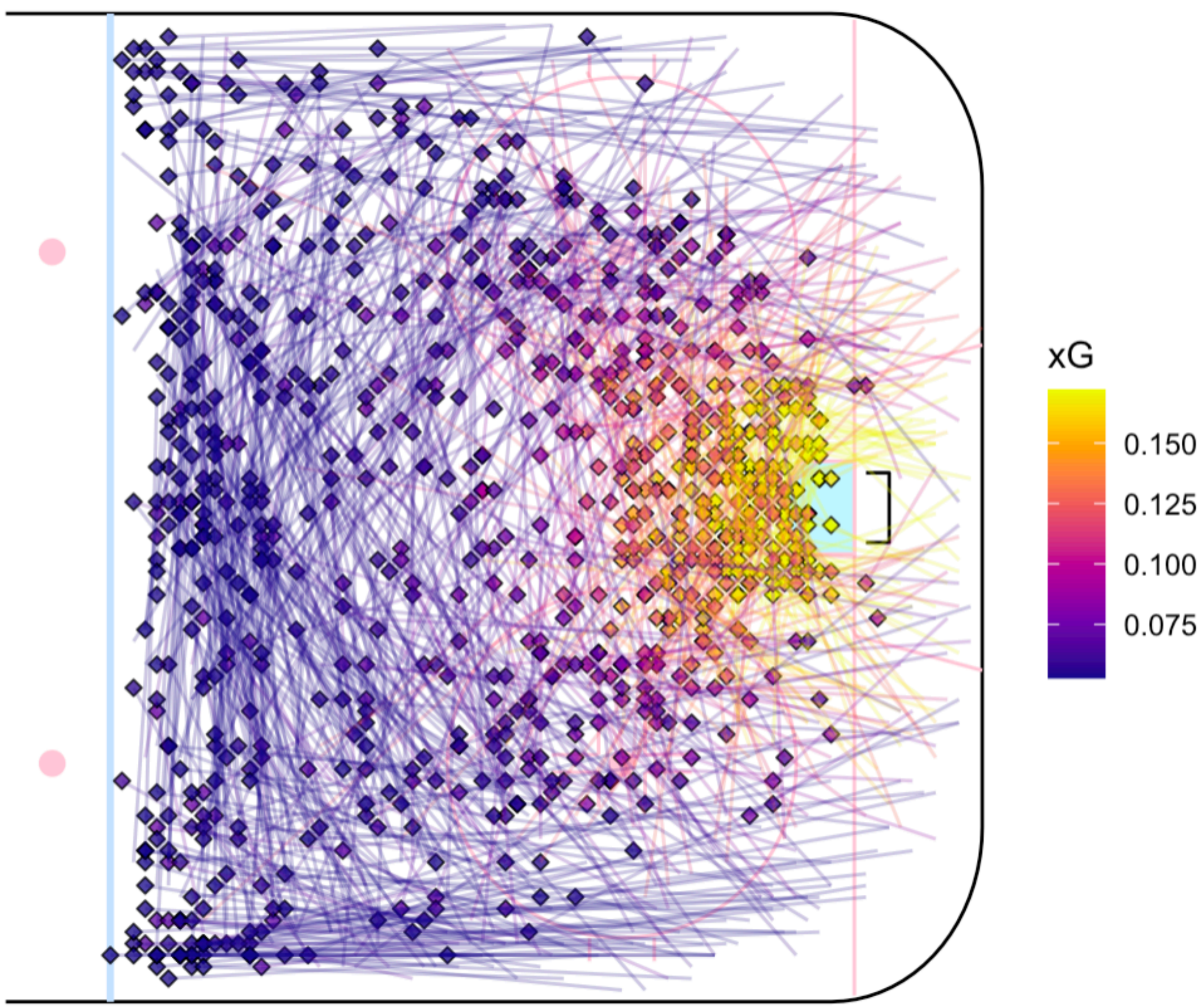

Figure 5: All one-timers in the dataset visualized

The two pass-related variables, pass distance and pass-shot angle difference, also revealed interesting patterns. The initial hypothesis proposed that pass-shot angle difference would be a highly important component of RRPs, given that passes that force goaltenders into large lateral movement are often cited as a reason for the success of RRPs. However, this hypothesis was not supported. The results suggest that the strength of these passes may not come from rendering the goaltender out of position and into large lateral movements, given the low percentage of variances explained by pass-shot angle differences. In fact, the successes of RRPs appear to be attributable to resulting shots that are close and well-centered, relative to the goal.

Additionally, the non-RRP passes in the top 10% of expected goals out-performed RPPs, even after adjusting for the most relevant variables (i.e., shot distance and shot angle) in predicting goals. These alternatives can be clustered into two groups. The first were “half” RRPs, where a pass crossed only one of the horizontal lines, resulting in a shot directly in front of goal. These passes are essentially RRPs in practice, and thus any team that is heavily defending against RRPs will likely have this guarded closely as well. The second identified RRP alternative were passes from behind goal-line beside the net, resulting in shots directly in front of the net. Although it is likely that a team defending close to the net (e.g., when guarding against RRPs) would also be able to defend this play due to the number of bodies in the area, it is a relatively understudied pass pattern that warrants further exploration.

In conclusion, the findings of this project are threefold. First, the results quantified the strength of the royal road pass relative to other one-time scoring chances. While there is a large variance in xG across one-timers, the xG of RRPs was 14% higher on average. Second, key contributors to the strength of RRPs are the proximity and central location of the resulting shots. This was also identified by Alex Novet (Novet, 2019), but was pushed to its logical limit by sampling only one-timers that are theoretically most affected by pre-shot movement. Third, passes originating from behind and beside the net also result in very high chances of scoring, as measured by this xG model. Although RRPs overall result in higher expected goals, shots in short distance and directly in front of the goal appear to have similar or better results. Therefore, it is likely that other passing areas and patterns may also result in high expected goal chances. Evaluations of pass/shot combinations should thus focus on shot-related variables (i.e., distance and angle); the rest (e.g., pass-related variables) may be a means to an end.

Future work

Although this study highlighted the importance of close shots, the position of the defense is what these shots hinge on, and how to open up opportunities to gain these chances. These topics could be examined using an expected threat model, much like what was presented as ISOLHAC 2020 by Sam Forstner (Forstner, 2020), and could also provide useful insight into defending these close shots. Additionally, the results indicate that passes from behind the net may be viable scoring chances. Future research should examine possible clustering of pass/shot combinations, similar to what David Yu presented at CBJHAC in 2020 (Yu, 2020).

Citations

Forstner, S. (2020, May 9). *Hockey Expected Threat (xT)*. Lecture presented at ISOLHAC 2020, Remote.

MacDonald, B. (2012, March 2). An expected goals model for evaluating NHL teams and players. Retrieved February 16, 2021, from http://hockeyanalytics.com/Research_files/NHL-Expected-Goals-Brian-Macdonald.pdf

Novet, A. (2019, August 16). Expected goals model with Pre-shot Movement, Part 1: The model. Retrieved January 29, 2021, from https://hockey-graphs.com/2019/08/12/expected-goals-model-with-pre-shot-movement-part-1-the-model/

Novet, A. (2019, August 16). Expected goals model with Pre-shot Movement, Part 2: Historic team and player results. Retrieved January 29, 2021, from https://hockey-graphs.com/2019/08/13/expected-goals-model-with-pre-shot-movement-part-2-historic-team-and-player-results/

Novet, A. (2019, August 16). Expected goals model with Pre-shot Movement, Part 3: 2018-2019 DATA. Retrieved January 29, 2021, from https://hockey-graphs.com/2019/08/14/expected-goals-model-with-pre-shot-movement-part-3-2018-2019-data/

Novet, A. (2019, August 16). Expected goals model with Pre-shot Movement, Part 4: Variable importance. Retrieved January 29, 2021, from https://hockey-graphs.com/2019/08/15/expected-goals-model-with-pre-shot-movement-part-4-variable-importance/#more-23593

Valiquette, S. (2015, March 6). The (royal) road to the future of goaltender analytics. Retrieved February 20, 2021, from https://www.msgnetworks.com/2015/03/06/the-royal-road-to-the-future-of-goaltender-analytics/

Younggren, J., & Younggren, L. (2018, June 7). A New Expected Goal Model for Predicting Goals in the NHL. Retrieved February 1, 2021, from https://rpubs.com/evolvingwild/395136/

Yu, D. (2020, February 8). *A Comprehensive Analysis of Pass Difficulty, Value, and Tendencies in Ice Hockey*. Lecture presented at CBJHAC in Nationwide Arena, Columbus, Ohio.

Code

See attached file name Geary_BDC21.rmd.