Passing patterns, passing efficiency, and Expected Goals on the Power Play

Ben Brill and Will Mosher
Big Data Cup - High School and Undergraduate category

The Setup

Whether you're a casual viewer or a lifelong hockey analyst, you know that efficient passing is key to a great offense. Cross crease, outlet, stretch or just a good ol' fashioned tape to tape, good passing leads to goals. While a gorgeous goal might make the highlight reel, a scoring play or a good look always starts with crisp, effective passing. For our question of interest, we wanted to focus on what types of passes resulted in the high efficiency and effectiveness. Specifically, which types of passes and how many of those passes resulted in the best scoring opportunities. And to measure the effectiveness of passing, we used expected goals per each shot as well as an adjusted effectiveness expected goals.

When you're in the offensive zone, the goal is to create the best chance possible with the smallest margin of error. We looked at what passes best suited a team in achieving that goal We identified two types of passes -- low-to-high and lateral -- that we thought would be common during the power play and would create high danger scoring opportunities.

Taking data from the 2018 Women's Olympic Hockey Tournament and the 2019-20 NWHL season, we looked at offensive zone possessions on the power play and at even strength to see what the most common passing patterns during the power play and to determine which of those passes produces the best chances and which of those passes produces the best chances *efficiently.*

Expected Goals model

Our expected goals model was developed based off of those created by EvolvingWild and MoneyPuck. We similarly used an Extreme Gradient Boosted model using the binary, linear regression solver to calculate our expected goals, or `xG` values. Taking advantage of the additional shot data contained within the data set (whether or not the shot was a one-timer, or whether there was traffic on the shot), our model used the following variables:

1. Time elapsed in period
2. Number of home skaters
3. Number of away skaters
4. X and Y coordinates of the shot location
5. Type of shot (Wrist, snap, slap, etc.)
6. Whether or not the shot was a one-timer
7. Whether or not there was traffic on the shot
8. Whether or not the shot was taken on the rush
9. Whether or not the shot was off of a rebound
10. The distance from the net of the shot location
11. The angle of the net of the shot location
12. The difference in time between the last event

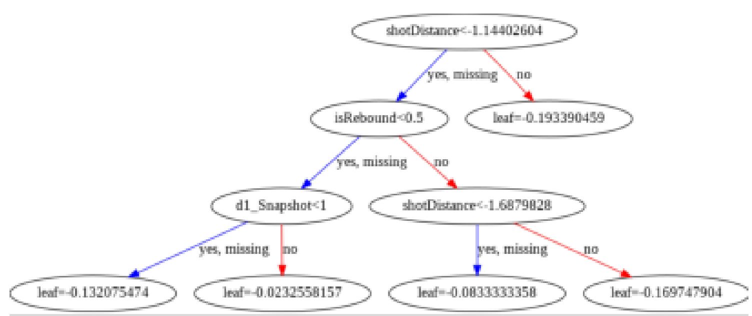


***Figure 1.** Coefficient decision tree generated by the extreme gradient boosted logistic regression model.*

Our model scored a 96% accuracy on our testing data. Overall, it accurately predicted a goal 91% of the time, the majority of which were **Type I** errors, or false negatives.

Passing Clusters

Now that we have a metric to evaluate how likely a goal is to be scored off of a given shot, we want to figure out how to classify each of the passes in the offensive zone. In the combined data set, there were 14,673 completed passes, which does not even include incomplete pass attempts. To try and find the most common patterns in the passing data, we used the KMeans unsupervised learning method to figure out the most common clusters of passes. We set `n_clusters` to 18, which means we wanted to summarize the 18 most common passes out of the 14,673 completed passes. We created patterns for both passes at even strength and on the power play.

The passing patterns at even strength and on the power play tend to be pretty similar. The two noticeable differences between two patterns are:

1. Passing at even strength appears to be more symmetrical, compared to Power Play passing which seems to favor the left side, probably due to a cycle running counter clockwise
2. There are more passes that originate from behind the net at even strength than on the power play.

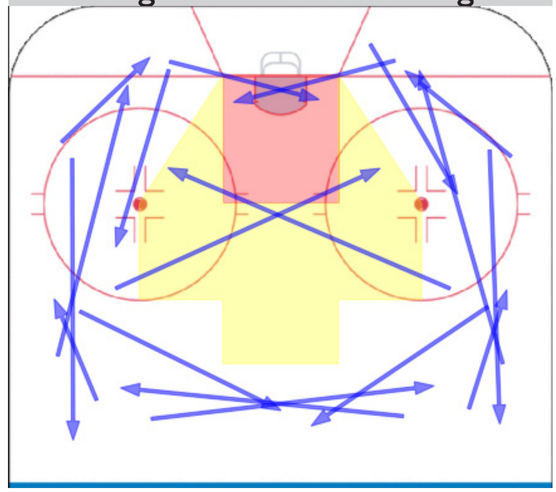
In addition to generating these clusters computationally, we classified them based on common pass names (i.e. behind the net, D-to-D, Umbrella-to-D, etc).

Lateral and Low-to-High passes

However, more importantly for our question, lateral and low-to-high passes are definitely among the most common passes, including on the power play. We clearly see in figure to the right that there are less lateral passes attempted than low to high passes. In addition the spread of the lateral passes is far greater than the spread of low to high passes.

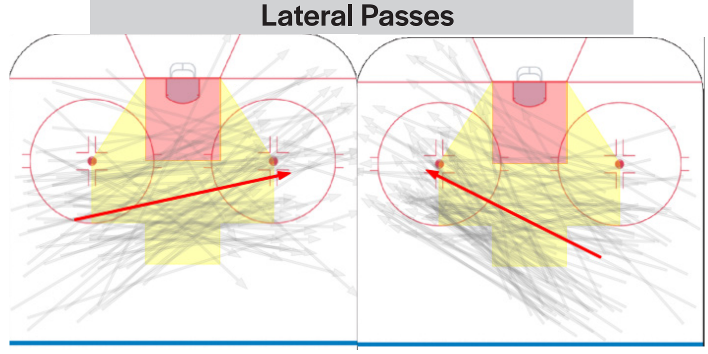
	Completed	Attempts	Completion %
Lateral	131	179	0.731
Low-to-High	123	115	0.911

Passing Clusters at Even Strength

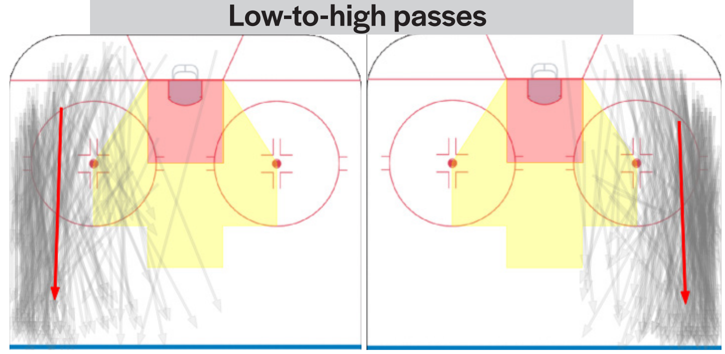


Passing Clusters on the Power Play

Lateral Passes



Low-to-high passes



Quantity of pass effect on Expected Goals

Not all passes are created equally, unfortunately. We now want to see the effect making a certain type of pass has on the `xG` of a shot off of a possession during the power play. In addition, we want to see if increasing the number of passes made -- regardless of their quality -- can result in a higher `xG` during a possession.

We defined a possession as the series of events that takes place from the time the team on the power play takes possession of the puck (regardless of the zone) to the time that same team either a). attempts a shot or b). loses possession of the puck. Under these rules, shot attempts directly off of rebounds count as their own possession.

In each of the possessions where a shot concludes the possession, we calculated how many of each of the 18 types of passes were taken before the shot, as well as the number of total passes, and the number of passes taken specifically in the offensive zone. We then analyzed the effects of the quantity of each of these types of passes and their effect on the shots they generated.

Quantity of total passes prior to shot attempt

We found that on the power play, there is a slight correlation between the number of passes made prior to the shot and the resulting `xG` value. Though there are a few fluctuations, the mean `xG` for each quantity of passes increase as you increase the pass number. The fluctuations are probably due in part to the smaller sample size for certain pass quantities. Though the max number of passes on a power play possession resulting in a shot attempt was 17, we limited our analysis to only look at the range [0,8] number of passes prior to a shot attempt.

It is important to keep in mind that the `xG` values with 0 passes prior do not include rebound chances, which we chose to exclude from our analysis. The mean `xG` for 0 pass attempts prior to a shot including rebounds was `0.062`.

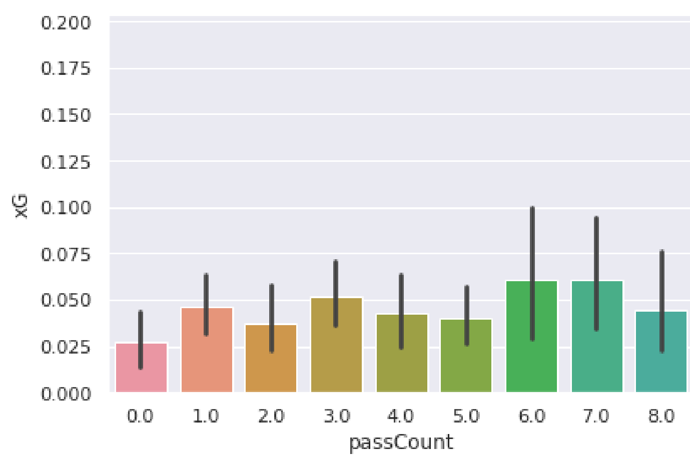


***Figure 3.** The mean xG per total passes completed per possession, with a bootstrapped confidence interval*

However, these pass attempts do include those generated off of the rush, which naturally will cause a higher `xG` than normal

Quantity of offensive zone passes prior to a shot attempt

An increase in the number of offensive zone passes before a shot attempt can similarly predict an increase in `xG`. These passes could also be more directly tied to resulting in a goal, as the total pass count can include passes made on the breakout or in the neutral zone.

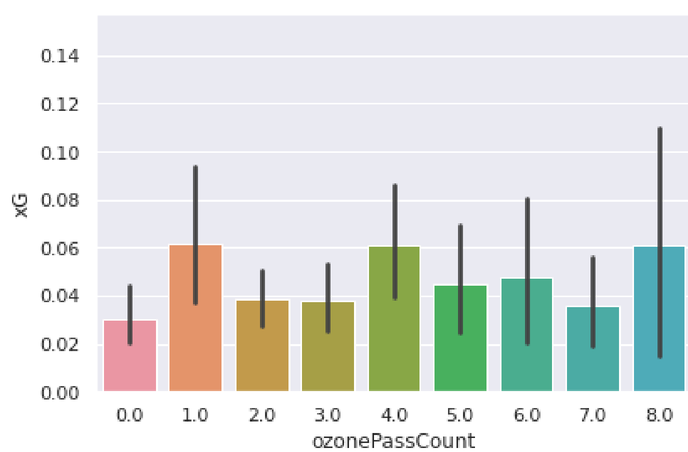


***Figure 4.** The mean xG per total offensive zone passes completed per possession, with a bootstrapped confidence interval*

Type of pass effect on Expected Goals

Again, not all passes are created equally. By looking at **xG** values by pass cluster, we can see there is a clear connection between the type of passes completed during a possession and the likelihood of generating a quality scoring opportunity and getting points on the board

Each of the 18 mean **xG** values for each pass cluster were calculated using shot attempts that stemmed from a possession that included at least one type of that pass. This means that number of possessions used to calculate all of the mean **xG** values is greater than the total amount of possessions, since some possessions will contain more than one pass type and will subsequently be used in the calculation of the mean **xG** of all the included pass types.

Let's take a look at our passes of interest: low-to-high and lateral passes

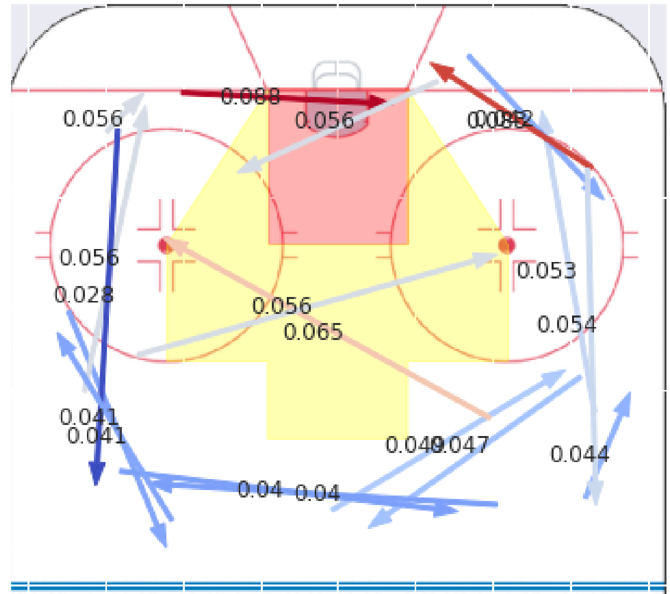


Figure 5. The mean xG per past cluster in the offensive zone on the power play. Each arrow represents the centroid of one pass cluster.

Low-to-High Passes vs. Lateral Passes

We previously hypothesized that low-to-high and lateral passing would be both two of the most common types of passes and two of the most effective passes in terms of creating scoring chances and goals.

As expected, lateral passes yield a higher average **xG** than low-to-high passes. While low-to-high passes create open space for the defensemen to get shots on net, the shots generated are typically coming from lower danger zones like the point.

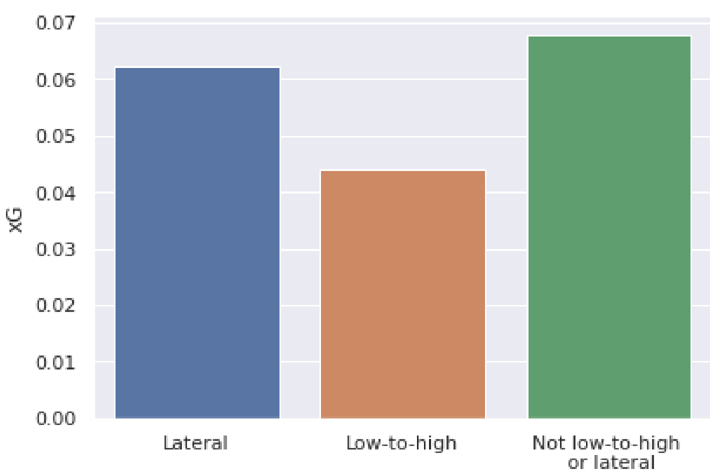


Figure 5. The mean xG for lateral, low-to-high, and neither pass clusters

In contrast, lateral passes, if successfully completed, disrupts the flow of the penalty killer's defense and places receiver in a high danger zone. However, as these zones are more carefully marked by penalty killers, it is harder for these passes to get through.

Using our completion rates for lateral and low-to-high passes we calculated before, we can calculate an "effective **xG**" in order to balance the trade-offs of each pass type. We will define this as:

$$Effective\ xG = xG_i \times completionRate_i$$

Pass	Completion Rate	xG	Effective xG
Lateral	.730	0.062	0.045
Low-to-high	.911	0.044	0.040

After adjusting for pass completion rate, the effective **xG** become very similar. By running a t-test with $\alpha = 0.05$ on both the original **xG** and the effective **xG** to determine if a significant difference exists between the two types, we find that a significant difference does exist between the original **xG**, but not the effective **xG**.

Comparing all different pass types

Since we determined that our two passes we were focusing on -- low-to-high and lateral passes -- result in no significant difference in `xG` when accounting for pass completion rate, let's apply this same logic to all of our pass clusters.

Of all of our pass types, behind the net passes have the highest effective `xG`. Passes out front had the highest original `xG` at `0.072`, but only about 58% of those pass attempts are completed. In contrast, passes behind the net originally had an `xG` of `0.066`, but those passes are completed about 77% of the time.

Though this is not what we expected, why might passes behind the net be the most effective at producing quality opportunities on the power play? Movement is key on any power play, so moving the puck behind the net will likely disrupt the penalty killers' formation. These passes in turn, can lead to a higher chance of executing some of the other types of high risk, high reward passes --especially passes out front -- due to the open ice created by the movement.

In fact, the appears to be a slight negative correlation between `xG` and the completion rate of a certain pass. This follows our logic of "high-risk, high-reward" passing. But then again, it appears that the "high-risk" balances out the "high-reward" in most situations.

Comparing to even strength

Though we looked primarily at the power play, the non-adjusted `xG` appear to be relatively similar at even strength as they are on the power play. In some cases, the values are slightly higher, but usually only by a few one-Thousandths.

Though we did not investigate this question fully, it remains to be seen whether or not the completion rate of these passes can counter-balance the potential high danger scoring opportunities. We focused primarily on the power play, because we believed that more open ice would lead to more completed passes, but we believe that the effective `xG` would be lower at even strength, due to less successful passing

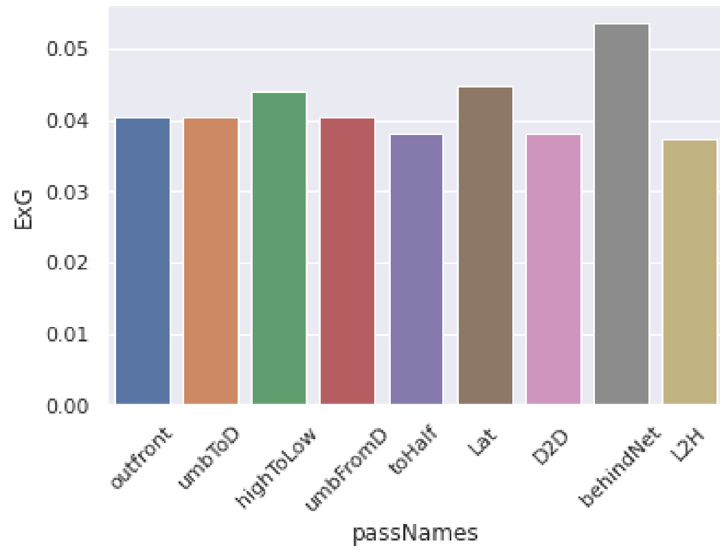


Figure 6. *The effective mean xG for all different types of pass clusters*

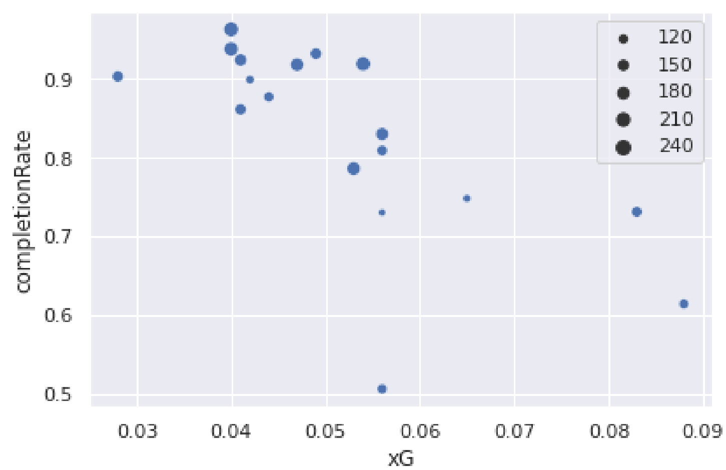


Figure 7. *xG plotted against pass completion rate, with the number of attempts indicated by the size of the point*

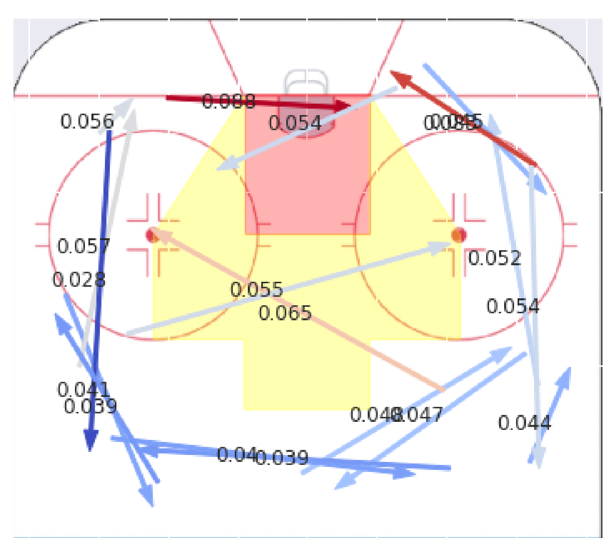
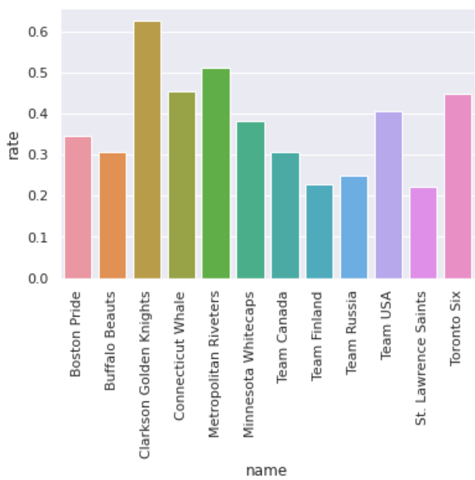


Figure 8. *The mean xG per past cluster in the offensive zone at even strength.*

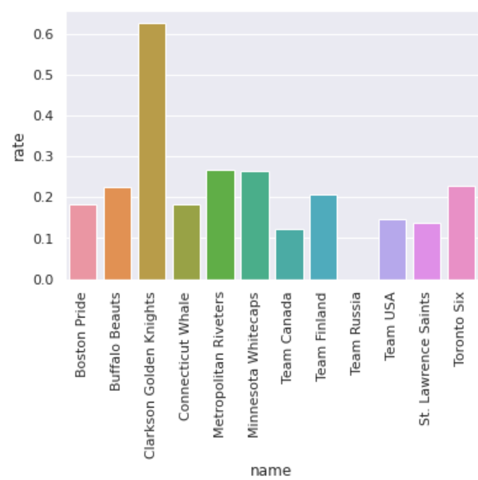
In Practice

Now that we see what teams should be doing, lets see what teams are actually doing in both the NWHL and during the Olympics:

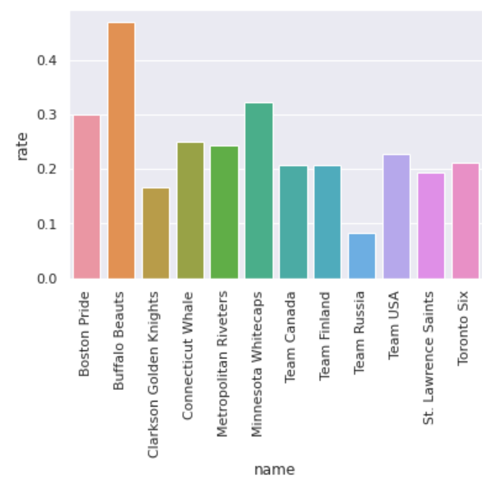
Percent of PP possessions with low-to-high passes



Percent of PP possessions with lateral passes



Percent of PP possessions with behind the net passes



The Clarkson Golden Knights really like to run their power plays using both the lateral and low-to-passes, a line of thinking we also held before running the numbers. Team Russia really did not want to be aggressive with their passing during their power plays, attempting a grand total of 0 lateral passes.

In contrast, the Buffalo Beauts attempt a less than average amount of lateral and low-to-high passes but attempt behind-the-net passes more than any other team. Team USA also attempted more behind the net passes than any other Olympic team.

Conclusion

After running the code, significance tests, and taking each scenario into context, we are able to delineate two things: Lateral passes are more effective while high low passes are more efficient. In context, this makes a lot of sense. Being able to whip the puck around from one point to the other makes the goalie have to move side to side (which is the most difficult move for goalies to make), as well as allowing time and space for players to set up shop around the net. Lateral passes create more motion, more motion creates more gaps and seams, more gaps and seams create more opportunities.

In contrast, passes up along the edges and behind the net don't exactly get you into a high danger scoring area, but are essential in the flow of good puck movement. The high effective `xG` for passes behind the net demonstrates this. Penalty killers and overall defensive strategies tend to clog up the middle while allowing the offensive team to swing the puck around the side and the perimeter with relative ease. While these passes themselves do not create the highest expected goals, they are essential in setting up a high danger scoring plays.

The main takeaways from our project is that all passes have their own strengths and weaknesses, but when you put them all together, while not overdoing it, you create opportunities. Focusing on working hard in front of the net, as well as throwing cross ice passes is the key to generating good looks. Rotating the puck from high to low sets up these opportunities for lateral passes. Solid passing leads to quality scoring chances from high danger areas. These are goal scoring plays. These are game winning plays.

Code can be found at https://github.com/benbrill/Big-Data-Cup-2021