Teamwork Makes the Dream Work - Determining Which Pass Types Create High Quality Chances in Women's Hockey

Ian Astalosh

1 Introduction

Hockey is a team sport, and the fastest way to move the puck up the ice has always been by passing between teammates. However, not all passes are created equal. Some passes keep possession, but don't do anything to increase a team's chances of scoring. Some passes are more to the point, and while have they have a great chance of leading to a shot if successful, also have a high chance of failing. The aim of this report is to determine which passes contribute the most to generating shots on goal, which famously, lead to actual goals. If certain passes or sequences were determined to provide a greater value, coaches could implement such strategies into their gameplanning and work on them more in practice to exploit that advantage.

The way we will do this is outlined below:

1. Group all available passes into pass clusters that can be compared and contrasted
2. Build an Expected Goals model specifically for the women's game to quantify the quality of shots
3. Develop a method of valuing individual passes in a given possession, which will then be used to assess the value of passes in each cluster

2 Methodology

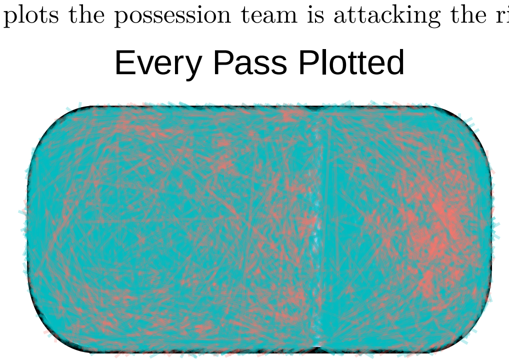
2.1 Data

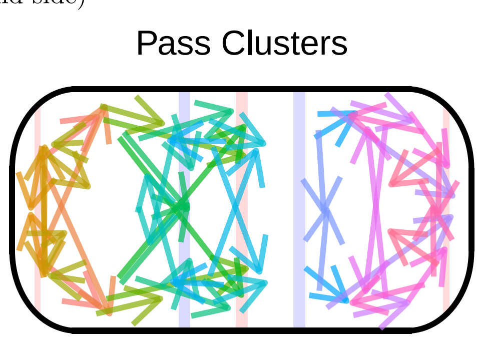
The data utilised is the combined sets of International, Collegiate and NWHL event data provided by Stathletes for this competition. Only events that took place with the teams at even strength (5 on 5) are considered, due to the woman advantage likely having a significant effect on the way the game plays out in that period. Examining the passing behaviour of both the power play and short-handed team could definitely be an interesting piece of research, but are beyond the scope of this report.

This leaves us with 38,060 events including 9972 plays and 5096 incomplete plays (15,096 total) as well as 2493 shots and 85 goals.

2.2 Categorizing Types of Passes

In order to determine which types of passes lead to scoring chances, we firstly need a way to distinguish between different pass types. Clustering the passes firstly simplifies the analysis, as instead of looking at 15,000 different passes, we can focus in on a much smaller number of categories of naturally similar passes, as seen below: (Note that in all plots the possession team is attacking the right hand side)



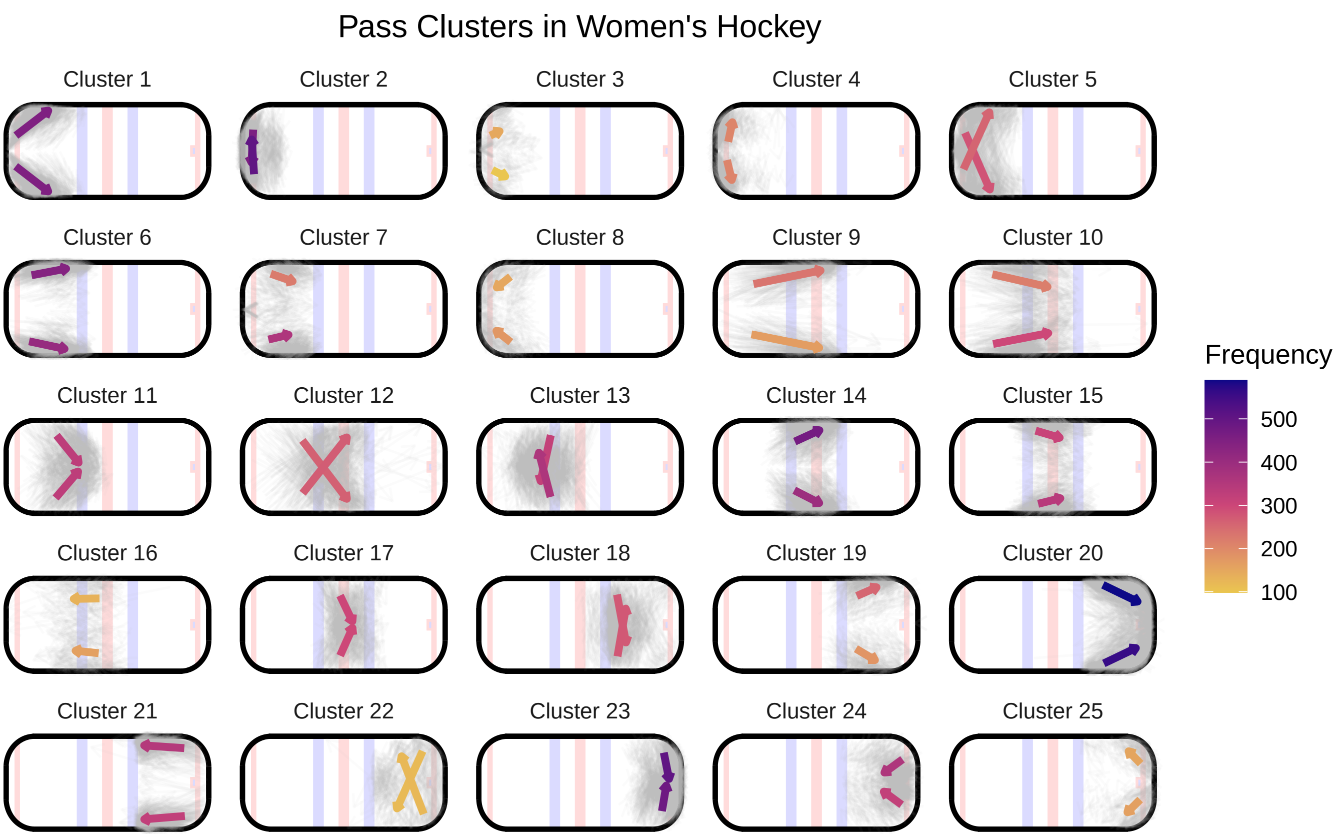


In order to generate these clusters, a k-means clustering algorithm is used. K-means essentially groups together similar observations based on their characteristics, which in our case here is the start and end locations of the pass, the pass length, the pass angle, and the distance from the centre of the ice. This type of clustering idea has been adopted several times for the men's game, in particular by Ryan Stimson and Daniel Weinberger on Hockey Graphs, and David Yu at the 2020 Columbus Blue Jackets Hockey Analytics Conference.

We largely follow Yu's approach to perform the clustering, which involves firstly mirroring all passes originating in the top half of the ice. Then, the k-means algorithm is performed, creating 25 clusters. The choice of 25 clusters (significantly less than the 100 Yu used) was based mostly on the sample size we had available as a number large enough to create distinctly different groups, but not so many the membership of those groups would be too small to do anything with. Unlike Yu, we do not unflip them after clustering to create an identical, symmetrical cluster. We assume that passes on the left side of the ice are not fundamentally different to their mirror images on the right, and so we treat them as part of the same category.

Below we display our 25 pass clusters, representing the 25 average types of passes in women's hockey. Each plot shows:

- The 'average pass' that the cluster represents. There are two arrows, highlighting the pass from both sides of the ice.
- All passes in the dataset that belong to that cluster as a shadow behind the average pass.
- The number of passes contained in that cluster. Some pass types are definitely more frequent than others!



We can see how our clusters roughly resemble the passes we see all the time in hockey games. Clusters 9 and 12 contains stretch passes, long breakout passes that aim to hit an attacker near the blue line. Cluster 21 contains cutbacks from the goal line to a waiting defenseman at the blue line. This is good, as the algorithm is able to identify passes commonly seen in game action. The most common passes on the ice originate from behind the possession team's own goal, as teams begin their attack (Clusters 1 and 2). The most common passes in the offensive zone are from the goal line straight back to the blue line, or behind the offensive goal (Clusters 21 and 23).

Examining these clusters is interesting on their own, as they can give us a feel for how teams or players tend to play. However, in this report we are not interested in tendencies, but rather how these passes increase the chances of scoring.

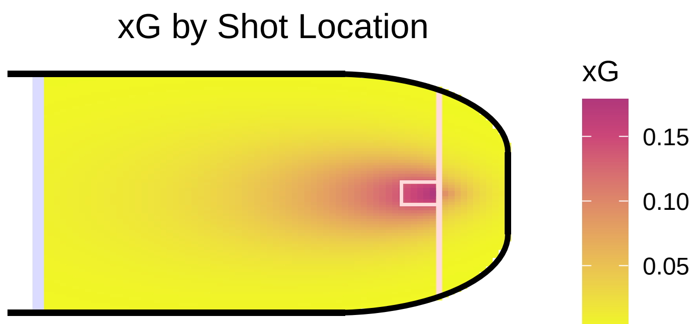
2.3 Building our xG Model

In order to evaluate which types of passes create better shots, we need a way of actually quantifying those shots. For this, we build a simple xG model utilising women's data. The model is a generalised linear model with a logistic link function, which generates the likelihood of that shot being a goal. The features in the model are:

- x coordinate of the shot
- y coordinate of the shot
- distance to goal
- angle in degrees from the shot location to the goal
- the time difference between the shot and the previous event in the data
- an indicator of whether the shot location was behind the goal line

A home team indicator was not added as it is assumed there was no home ice advantage due to all international games taking place in South Korea, and all NWHL games taking place in a bubble with no live fans.

The below plot shows the expected goals for shots within the offensive zone, and these values are used in the next section to evaluate the quality of each chance.



2.4 A Novel Method of Quantifying Pass Value

While more advanced models such as xThreat have been proposed to quantify the impact of individual events, we utilise a simpler approach to determine which passes add expected goals. The approach is inspired Statsbomb's xGChain and xGBuildup, which isolates possessions into the events that comprise the possession. However, instead of then attributing any expected goals to all players involved, we attribute a weighted amount of expected goals to the passes involved, as we are interested in their value.

The approach can be broken into several parts:

1. Define what constitutes a possession.

I define a possession as the time between when a team obtains possession of the puck, to when they either lose it or deliberately cede it (ie. when a team takes a shot. If a team takes a shot but gathers the rebound, this should count as a new possession, because they had already chosen to end it by taking the shot. No one takes a shot hoping it won't go in). An incomplete play that was recovered by the possessing team I counted as retaining possession, however.

In short:

- *Possession starting events:* Faceoff wins, takeaways, puck recoveries following a shot
- *Possession ending events:* Takeaways, penalties, incomplete passes (that were recovered by opposing team), shots, goals.

2. Assign a value to that possession. We create $PossessionXG$ which is simply the expected goals of any shot that took place on that possession (note by definition, a possession can only have a maximum of 1 shot). Possessions that do not contain a shot are assigned a value of 0.

3. Allocate a proportion of the possession value to each pass based on how close they were to the shot. Individual passes within a possession are assigned a proportion of the total possession value based on how close they were to the shot taking place.The first pass of a 6 pass chain is important, but not as important as the pass immediately before the shot. The formula use to distribute the total possession value to each pass is given by:

$$IndividualPassValue = PossessionXG \times \frac{13 - n}{13}$$

where n is the number of passes between the current pass and the shot (ie. $n = 0$ for the shot assist). 13 was chosen as the scaling factor because that was the most passes recorded on a single possession (a 41 second masterclass by the Toronto Six against the Minnesota Whitecaps that promptly ended in an… incomplete pass).

So, perhaps we had a particular sequence that was a Cluster 20 pass, followed by a Cluster 23 pass, followed by a Cluster 24 pass, which resulted in a shot with an xG value of 0.1. Then the value attributed to each pass would be:

- Cluster 18 (the assist): 0.1
- Cluster 23 (the pass to assist): 0.092
- Cluster 20 (the pass to pass to assist): 0.085

4. Average over all clusters to determine the clusters average value in build up play Once we have assigned values to each individual pass on each possession, we can average over all passes to determine which add the most value. ie.

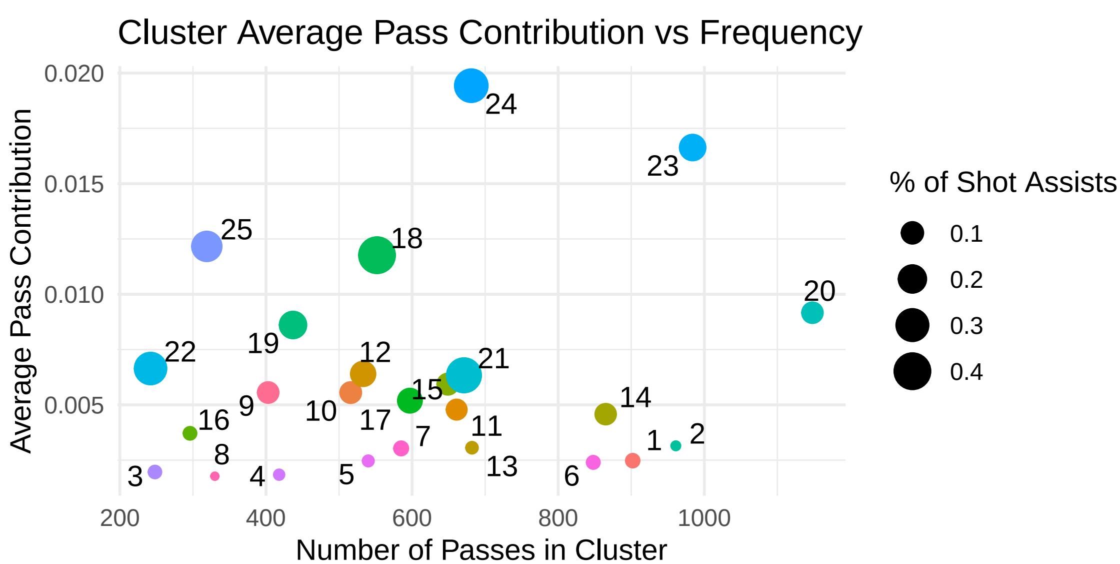
$$AvgPassContribution = \frac{\sum^{cluster} IndividualPassValue}{n_{cluster}}$$

where $n_{cluster}$ is the total number of passes categorized as having that cluster.

This Cluster process will enable us to obtain a single value for each pass cluster, representing the contribution of that pass type to creating chances.

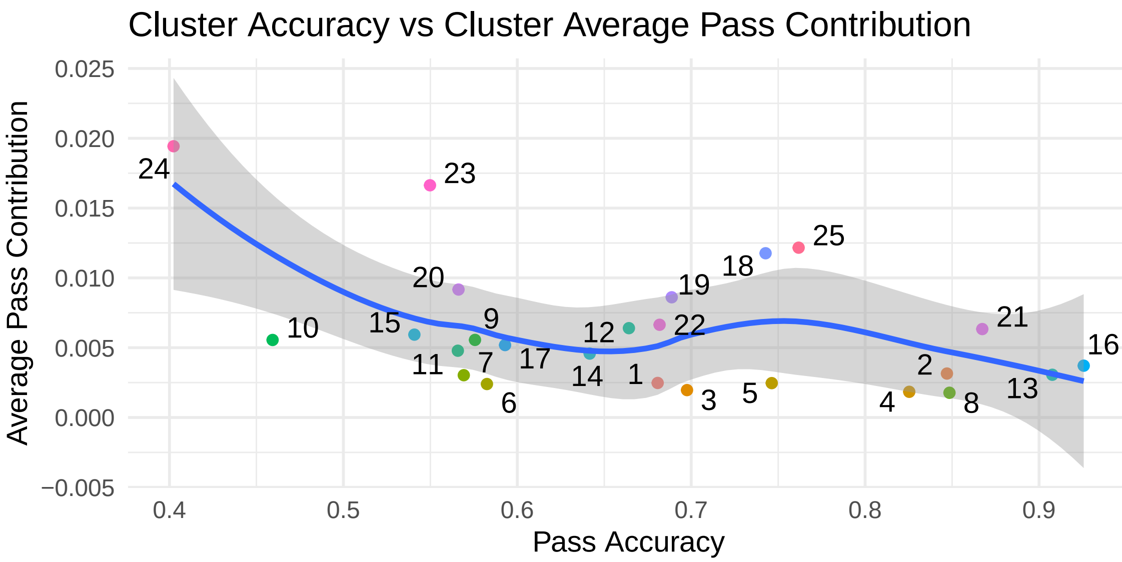
3 Results

We now put this into practice. We have 11,588 distinct match possessions available in our dataset, of which 2578 produced a shot.



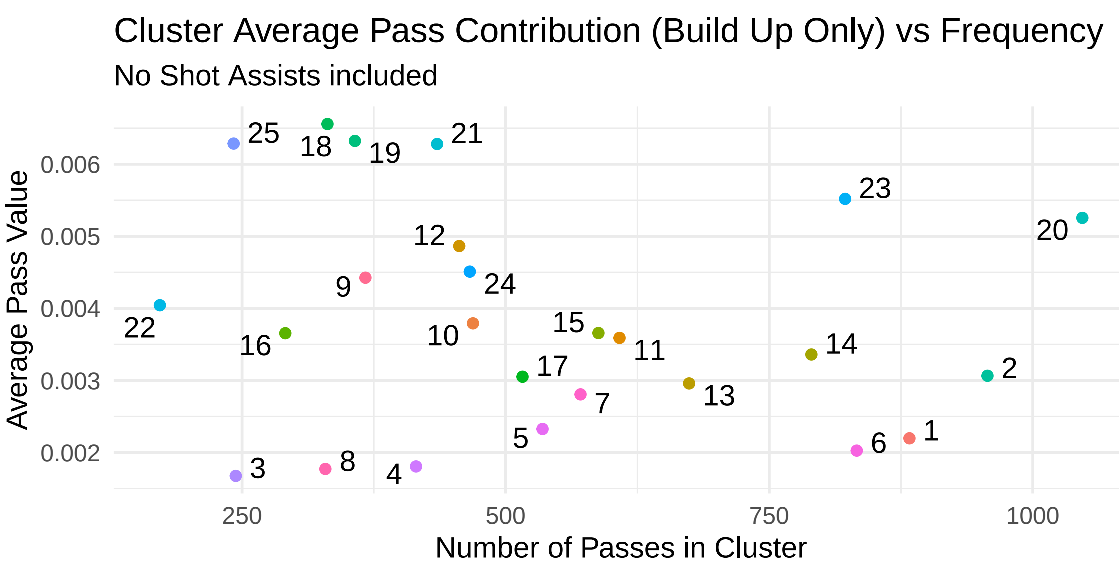
Unsurprisingly, the pass types that have the greatest value are the ones that occur in the offensive zone (Clusters 18 onwards). The highest valued passes in Clusters 23 and 24 direct the puck towards the goal, so it makes sense that they add the most value as they lead to the most dangerous shots. Somewhat surprisingly, the largest cluster, Cluster 20 (the pass from the goal line back to the blue line) has a comparatively low value per pass. This also makes sense as the puck goes further from the goal, but perhaps teams when stuck in the corner should look to fire to the middle or around the back rather than kicking it out.

One thing to note here is the relationship between the accuracy of passes and value added.



Just because a pass is a high percentage pass, and means you'll keep the puck, does not appear to mean it actually increases your chances of scoring a goal. The highest reward passes are also the highest risk, but teams could be encouraged to be more aggressive with the knowledge that a slight increase in goals could follow.

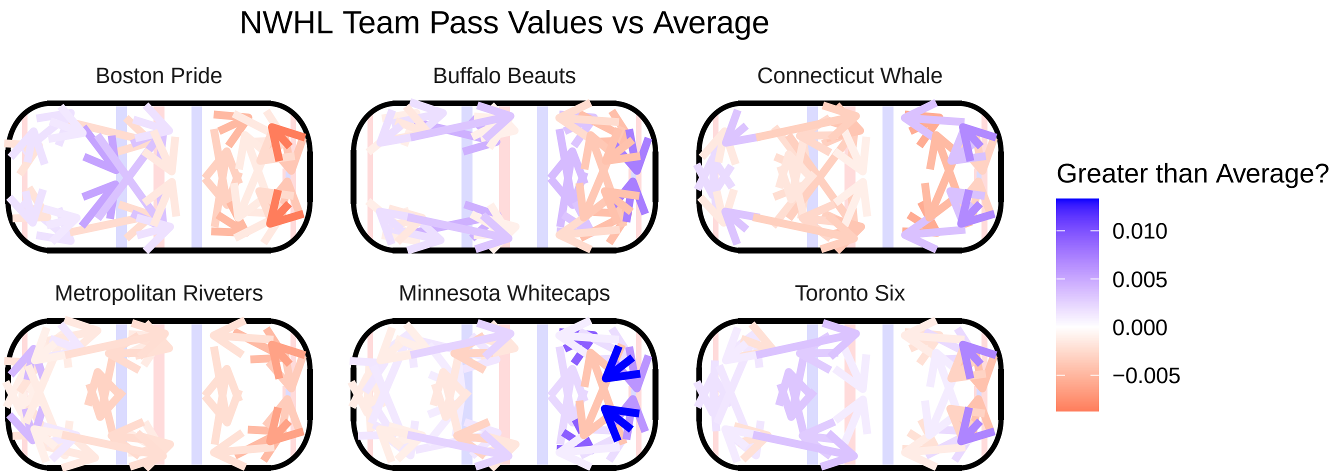
From the original graph, it is clear the clusters that add the most value are the ones that have high shot assist percentages. This is, quite often they are the last pass before a shot is taken. If we want to look purely at the contribution in the build up to shots, we can remove all these shot assists and look only at the contribution of passes before that:



With the shot assists removed, we can see how certain passes originating from the defensive and neutral zones are useful. Although the passes in the offensive zone still have the greatest value towards scoring, Clusters 9 and 12, stretch passes and breakout passes, also provide a significant amount of value, despite being smaller in their cluster size. This suggests that teams should longer, quicker passes when breaking out from their own end.

3.1 Team Performance

Now with a measure on the value of each pass type, we can group the data by team to see which teams got more and less value out of certain pass types than the average. Certain teams may excel at particular pass types due to their practice or particular strategies they are implementing. The below plots show which pass clusters that each NWHL team during the 2021 season either outperformed or underperformed their expected values.



Teams like Boston and Toronto appeared to get slightly more value towards shots on passes building out from the back. Minnesota appears to have increased value on passes in the offensive zone. Tools like this could be used by teams not just to see what particular teams like to employ when they attack, but also *what they are actually good at*.

4 Conclusion

The purpose of this report was to classify and analyse passes in women's hockey, to determine which passes are played most often and explore how they create chances. Creating our 25 pass clusters provides a framework to do this, showing that passes in the offensive zone that get the puck as close to the goal as possible add more value than passes that maintain possession. These passes may be high risk as they have lower accuracy, but if teams want to maximize their chances on goal, they appear to be risks they should take. Using this work, coaches could create better tactical gameplans on how to attack, focussing on the passing lanes which provide the highest chance of scoring. With more time and data, looking at the effect of particular passing sequences (ie. two or three pass combinations) would be a natural follow up, as well as incorporating this with tracking data and determining the most valuable passes when the defense is configured in a particular way.

5 References

1. Stimson, R 2017, *Identifying Player Styles With Clustering*, Hockey Graphs, viewed 5 March 2021, https://hockey-graphs.com/2017/04/04/identifying-player-types-with-clustering/
2. Weinberger, D 2019, *Visualizing and Quantifying Passing on the Power Play*, Hockey Graphs, viewed 4 March 2021, https://hockey-graphs.com/2019/09/04/visualizing-and-quantifying-passing-on-the-power-play/#more-23791
3. Yu, D & Columbus Blue Jackets 2020, *CBJ 2020 Hockey Analytics Conference*, viewed 5 March 2021, https://www.youtube.com/watch?v=TENd93QJt-8&t=7924s
4. Lawrence, T 2018, *Introducing xGChain and xGBuildup*, viewed 5 March 2021, https://statsbomb.com/2018/08/introducing-xgchain-and-xgbuildup/