Summer 2022

The Impact of Amateur Hockey Leagues on NHL Performance

Leah Emily Smith

The Impact of Amateur Hockey Leagues on NHL Performance

Leah Weak, Advisor: Courtney Thatcher
University of Puget Sound

Background

One of the main ways a National Hockey League (NHL) team can acquire players is through the Entry-Level Draft, where players in between the ages of 18 and 20 are selected by teams based on what they have accomplished in junior hockey. Over 200 players are drafted each year over the course of seven rounds. However, the chances of draftees making it in the NHL is low. Only 51% of players drafted from 2007-2015 have played in the NHL. And only 33% have played in at least 82 games. Since the success rate is so low it is important to understand what amateur leagues can tell us about a player and how they might contribute to their NHL team.

NHL. 2020

Questions

- What is the relationship between amateur hockey leagues and success in the NHL?
- What can we determine about what types of players amateur leagues in certain countries are developing?

Data

- Skaters (Forwards and Defensemen) drafted between 2007 and 2015.
- NHL statistics were taken from the 2007-08 season up to the 2021-22 season.
- All amateur league statistics used are from the season right before the player was drafted.
  - Undrafted players: season before first year of draft eligibility.

Measuring Success

- **Cap Hit %**
  - Salary-cap hit percentage of the upper limit at the time of signing contract.
  - The upper limit for the 2022-2023 season is $82.5 million.
- **GAR: Goals Above Replacement**
  - Attempt to capture total value of player in one number.
  - Represents the number of goals that a player contribute relative to a replacement-level player (GAR = 0).
  - Player value is separated into different categories and value is calculated in each, usually by a regression-based technique, and summed up for each player.

ANOVA/ANCOVA

- **Does the amateur league a player is drafted from affect their success?**
  - One-Way ANOVA: single categorical variable with multiple levels and numeric dependent variable.
  - Explore the relationship between amateur league and a chosen dependent variable.
    - *GAR ~ Amateur League*
  - There were no statistically significant differences between group means.
- **What is the relationship between GAR and amateur league while factoring in amateur league success?**
  - One-Way ANCOVA: two numeric variables and one categorical.
    - *GAR ~ Points per Game in Amateur League + Amateur League*
  - Points Per Game in Amateur League is significant, but amateur league grouping is not.

Logistic Regression

- **What is the probability that a North American forward will reach a certain level of success based on what amateur league they played in and how many points per game they scored while in that league?**
- Here we are looking at two levels of success:
  - **Figure 1:** Maximum Cap Hit% (over entire career) is in the 90th percentile (>0.067). In 2022-23 season, a cap hit % of 6.7% is about $5.5 million.
  - **Figure 2:** At least 1/3 of seasons played are considered successful.
    - A successful season is when a player is in the 75th percentile based on GAR (>5.7).
- While the two logistic regression models below were found to be statistically significant, there was little difference between the bigger leagues (OHL, WHL, QMJHL, USHL) but Other, NCAA, and High School are somewhat different. In particular, a forward in High School would have to score a lot more points to reach the same probability of success as the other amateur leagues for both measures if success. Note as well, that the Other group includes many different smaller leagues that have a smaller number of players in the NHL which may introduce some inconsistencies.
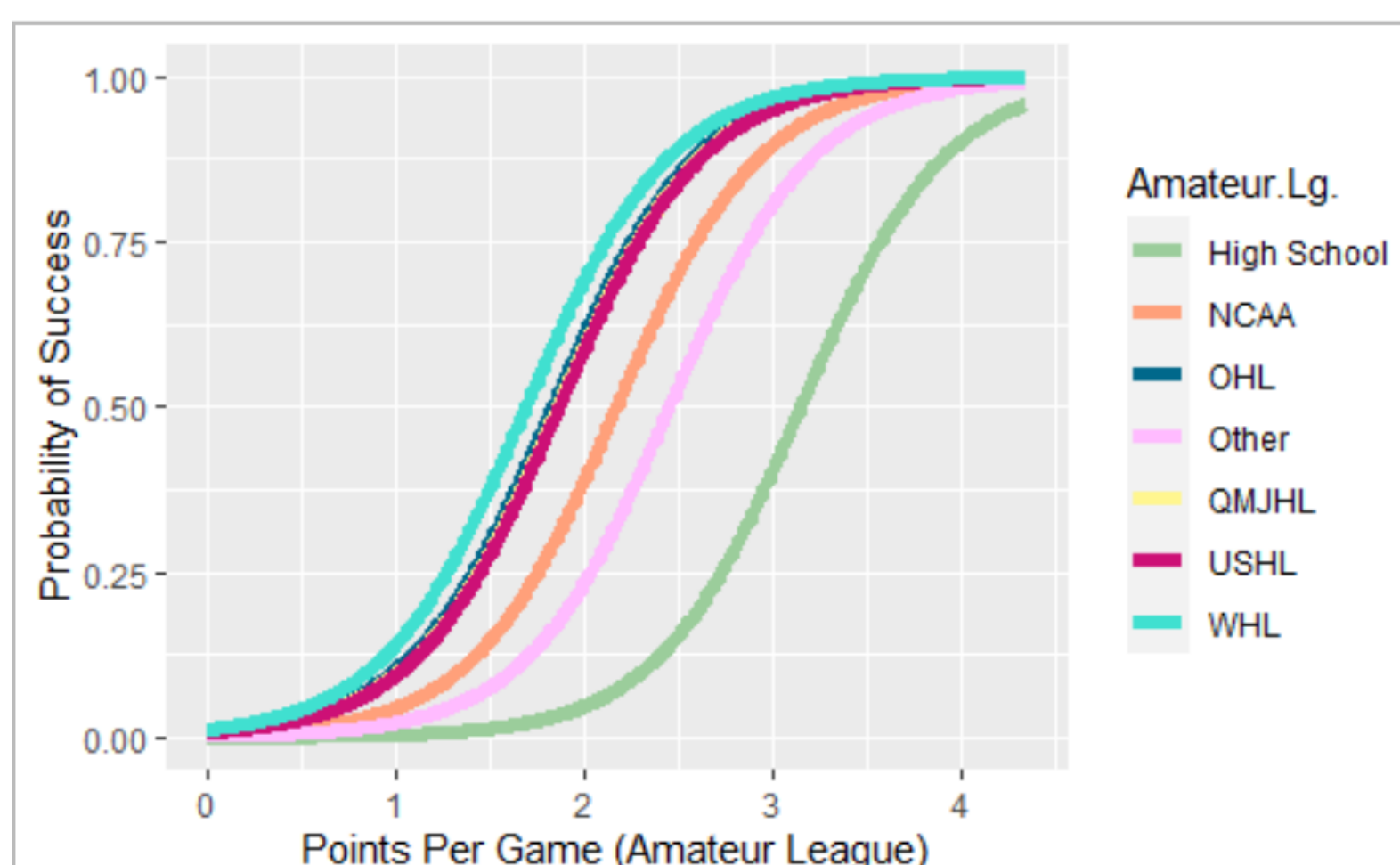
Figure 1. Plot of probabilities calculated by logistic regression. Each line shows the probability that a forward from that amateur league will be successful based on their max cap hit%.
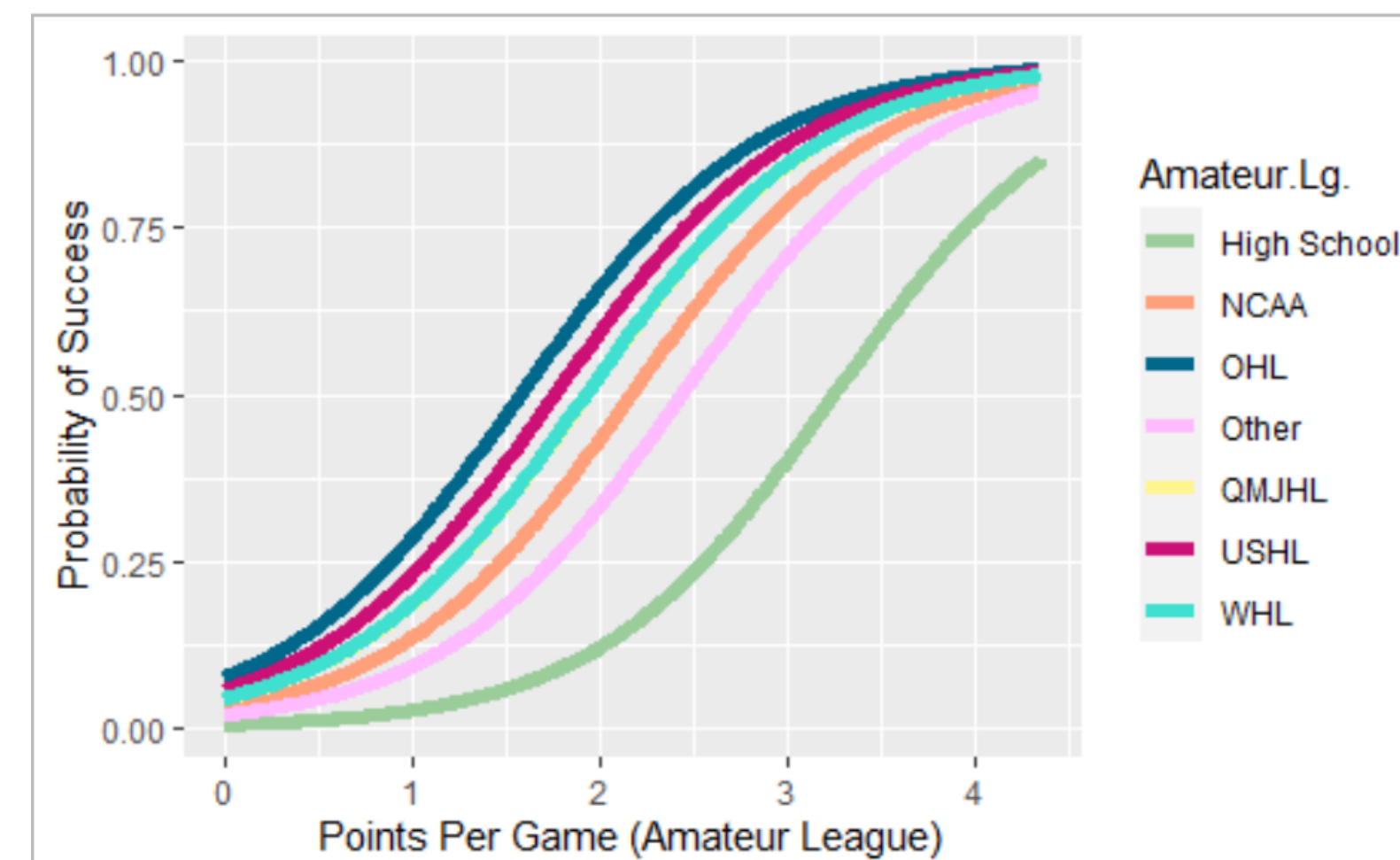
Figure 2. Plot of probabilities calculated by logistic regression. Each line shows the probability that a forward from the associated amateur league will be considered successful based on their GAR each season.

Hierarchical Clustering

- **Are certain countries better at developing a certain "type" of player?**
  - In this case, country does not necessarily refer to a player's nationality but instead where the amateur league they participated in is located.
- We use Hierarchical Clustering to group all players (forwards and defensemen) into clusters (types of player) based on 22 different normalized variables including goals, assists and time on ice per game (even strength, power play, shorthanded) as well as hits per game and zone entry percentages.
  - Note that only players who have played at least a full regular season (82 games) were included and every statistic is calculated for each season and then averaged based on seasons played.
- **Figure 3** shows an example of how clusters were classified, and **Table 1** shows the classification for all clusters.
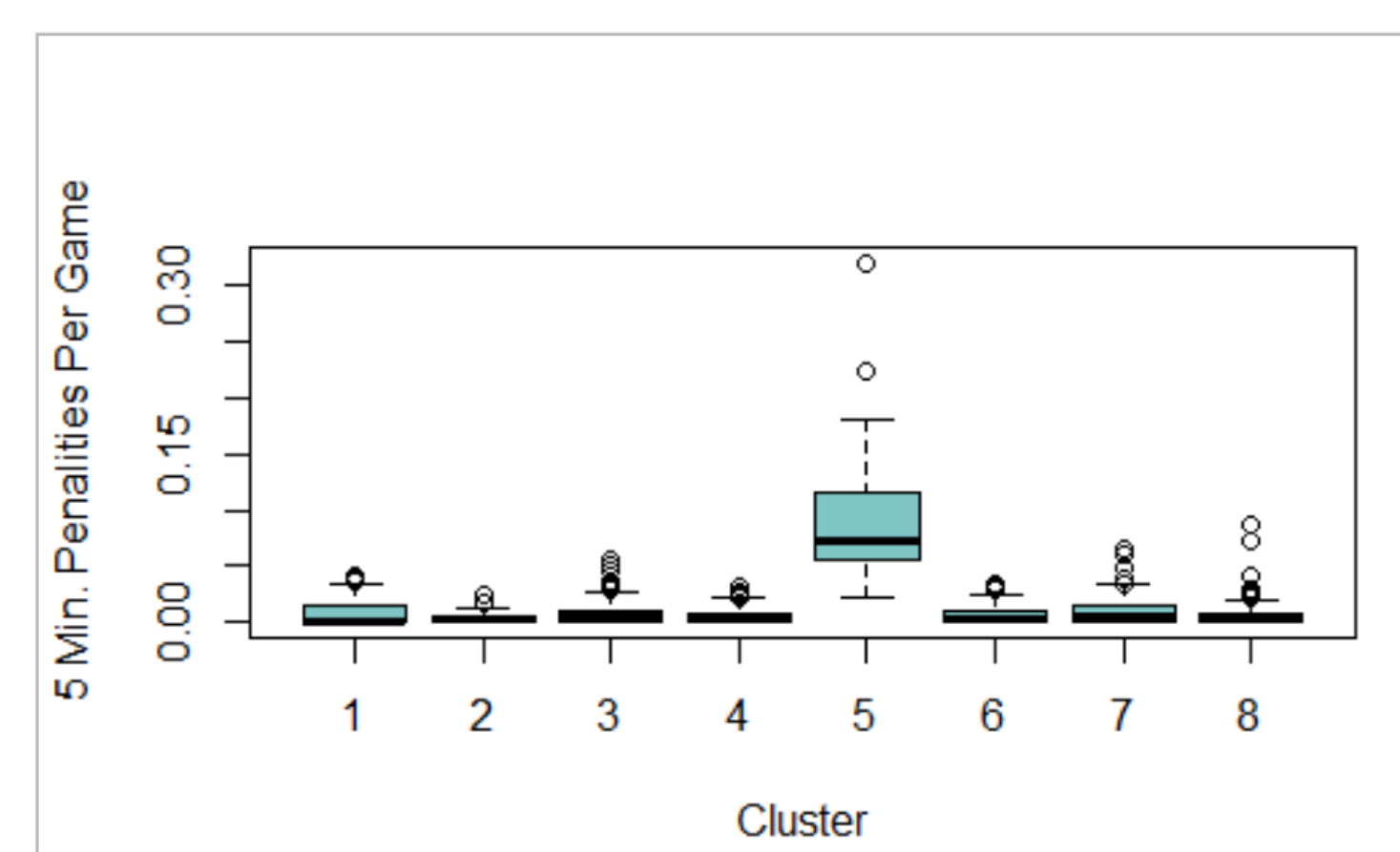
Figure 3. Boxplot of average five-minute penalties taken per game by cluster. Players in Cluster 5 have a higher value. Looking at this and the other boxplots of each statistic used (not shown), we can deduce that this cluster represents Enforcer type players.

Cluster	Classification
1	Replacement level player (any position)
2	Offensive Defenseman
3	Defensive Forward
4	Solid 2-Way Defenseman
5	Enforcer
6	Top Tier Forward
7	Defensive Defenseman
8	2nd Tier Forward

Table 1. Types of players represented by cluster.

Figure 4 shows what percent of players from each country fit into each cluster. Countries with a smaller number of players were grouped into Other (Europe). While there would need to be more work done to confirm these results there are a few interesting things to note.
- The heat map seems to indicate that most Enforcer type players (Cluster 5) are coming from Canadian and American amateur leagues.
- Sweden seems to be developing more players in Clusters 3, 4, and 8.
- Russia does not have any players in Cluster 2 indicating that they do not put an emphasis on developing defensemen who are more offensively minded.
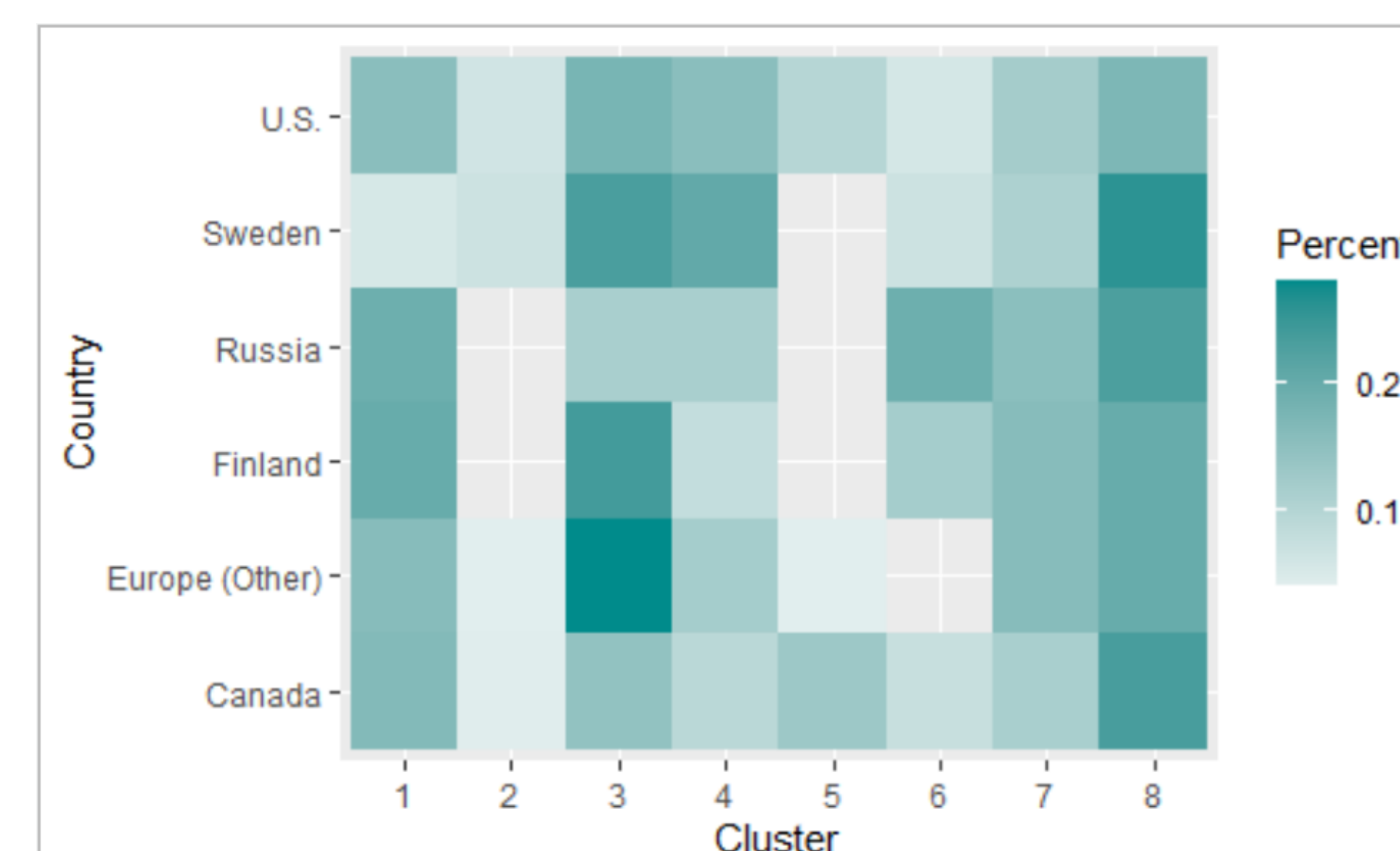
Figure 4. Heat map of percent of players in each cluster by country amateur league.

Future Directions

- Expand logistic regression to include non-North American amateur leagues and try different levels of success.
- Data collected for amateur leagues only included the year before a player was drafted or the year before eligibility for undrafted players. It would be interesting to track a player's development farther back.

Click On Detroit. 2021

Mile High Sports. 2022

Data Sources

CapFriendly. *Active Players* [date file].
Evolving Hockey. *Skater Tables* [data file].
Evolving Hockey. *GAR Skater Tables* [data file].
Hockey Reference. *NHL Draft Finder* [data file].
Ontario Hockey League. Player Statistics [data file].
Quebec Major Junior Hockey League. Player Statistics [data file].
United States Hockey League. Player Statistics [data file].
Western Hockey League. Player Statistics [data file].