

Quantifying the Contribution of NHL Player Types to Team Performance

Timothy C. Y. Chan, Justin A. Cho, David C. Novati

Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada
{tcychan@mie.utoronto.ca, justin.cho@utoronto.ca, david.novati@utoronto.ca}

In this paper, we use k -means clustering to define distinct player types for each of the three positions on a National Hockey League (NHL) team and then use regression to determine a quantitative relationship between team performance and the player types identified in the clustering. Using NHL regular-season data from 2005–2010, we identify four forward types, four defensemen types, and three goalie types. Goalies tend to contribute the most to team performance, followed by forwards and then defensemen. We also show that once we account for salary cap and playing-time information, the value of different player types may become similar. Lastly, we illustrate how to use the regression results to analyze trades and their impact on team performance.

Key words: hockey; NHL; sports applications; team performance; player types; k -means clustering; regression.

History: This paper was refereed.

Introduction

In professional sports leagues such as the National Hockey League (NHL), a team's general manager (GM) is tasked with assembling a winning team. In the NHL, a GM's job is complicated by the presence of a hard salary cap, which limits how much the team can spend on its players. Because the salary cap aims to facilitate parity across the league, GMs must build well-balanced, high-performing teams by leveraging any competitive advantage they can find. To make informed decisions, these decision makers must be able to measure the performance of different players and their contribution to the team's overall performance. As a first step, identifying distinct player types and quantifying the contribution of these player types to team performance would provide them with a framework to evaluate the value of different players.

Players are usually differentiated by their individual performance. In hockey, scientifically analyzing individual player performance is a challenging task because the game is so fluid. However, researchers have attempted to classify NHL players into different player types as a way to compare them and to analyze differences in salaries (Vincent and Eastman 2009; Jones et al. 1997, 1999). In this paper, we build on prior research by developing a novel classification

scheme that is well suited to current NHL players. Using this scheme, we measure the contribution of different player types to the number of team points a team achieves over the course of an NHL season. We then demonstrate how this knowledge can be used to evaluate trades and other roster changes.

Operations research (OR)-focused research in hockey has considered topics such as modeling the goal-scoring process (Mullet 1977), scheduling (Costa 1995, Ferland and Fleurent 1991, Fleurent and Ferland 1993), calculating the probability of making the playoffs (Ingolfsson 2004, Russell and vanBeek 2009), modeling different game states to determine successful team strategies (Thomas 2006), and modeling the win-loss percentage using a function of goals scored or allowed (Cochran and Blackstock 2009). Ingolfsson (2010) presents a comprehensive review of OR analyses in hockey.

Vincent and Eastman (2009) inspired our research. They proposed a classification scheme for NHL hockey players by clustering them using their career statistics up to and including the 2002–2003 season. They then identified major differences in player earnings between the clusters. Although similar in spirit, the work we present in this paper differs from the previous research in the areas of data, clustering, and objective.

Data

- **Post-lockout statistics:** In response to the lockout that caused the cancellation of the 2004–2005 season, the NHL instituted many rule changes to increase offensive production. (Note that the 2004–2005 NHL season was officially cancelled on February 16, 2005 because of a lockout. After the previous collective bargaining agreement expired on September 15, 2004, the NHL and NHL Players Association were unable to come to an agreement until the following summer). Because of these changes, the game play has changed considerably; therefore, pre- and post-lockout statistics are not directly comparable. By basing our analysis on post-lockout statistics, we believe that our clusters more closely reflect current player types.

- **Single-season statistics:** We focus on single-season statistics because players might change the role they play over the course of their career; ultimately, we are interested in relating a team's performance to the player types on that team during a particular season.

- **Additional statistics:** We include blocked shots and hits, which had not been considered previously (these statistics were not available on an individual basis prior to the lockout).

- **Salary cap data:** Instead of player earnings, we examine player salary cap contributions, which were not considered previously (because the salary cap was introduced after the lockout).

Clustering

- **New player types:** We put additional emphasis on defensive categories for forwards and defensemen, which resulted in novel player types.

- **Goalies:** Goalie clustering is new, and its inclusion allows us to more exhaustively attribute team performance to the players on the team. For example, a superstar goalie might account for higher performance than otherwise would be expected by considering forwards and defensemen alone.

Objective

- **Using player types to measure team performance:** Generating player-type clusters is only the first step of our analysis. The larger issue in which we are interested is the impact of different player types on the overall performance of a team.

The contributions we present in this paper are (1) a novel classification scheme for NHL player types in the post-lockout era, (2) a quantitative measure of the impact of different player types on team performance, (3) an analysis of player types to quantify the value of their contribution to team performance when accounting for salary information and playing time, and (4) a trade tool that can be used to estimate the impact of trades on team performance.

Hockey Background

An NHL hockey team typically dresses 20 players for each game: 12 forwards, 6 defensemen, and 2 goalies. Forwards are usually partitioned into four lines of three forwards each (one center, one left wing, one right wing). Defensemen are paired off in three pairs. Of the two goalies, one starts and typically plays the entire game, whereas the other is a backup and sits in reserve in the event the starter is injured or plays poorly. The standard number of players that can be on the ice at the same time from one team is six: three forwards, two defensemen, and one goalie. Players on the ice may switch with those on the bench during certain stoppages in play or anytime that the puck is live (as long as no more than six players of one team are on the ice at any time). That players may move on and off the ice while the puck is in play adds to the fluidity of the game. Typically, an entire line (three forwards) will go off at once and a new line will replace it. The same is true for defense pairings. A team can also exchange its goalie for an extra forward or defenseman at any time a legal substitution can be made. This is often done during the final minutes of a game by a team that is trailing by one or two goals. The question of exactly when a goalie should be pulled is a popular one among fans and academics (Morrison 1976, Morrison and Wheat 1986, Erkut 1987, Nydick and Weiss 1989, Washburn 1991, Beaudoin and Swartz 2010).

An NHL hockey game is played over three periods, each lasting 20 minutes. If the score is tied at the end of the third period, a sudden-death overtime period is played for up to five minutes. If the score is still tied at the end of the overtime period, the outcome is decided by a shoot-out. In a shoot-out, players try to score in a one-on-one situation against the

other team's goalie. Initially, three shooters from each team are scheduled, and shooters alternate by team. If a tie remains after the three initial shooters from both teams have finished, the shoot-out continues until a round in which one team scores (the winning team) and the other team does not. The winning team always receives two points in the standings (regardless of whether the game is won in regulation time, in overtime, or in a shoot-out), and the losing team receives either one point (if the loss occurs in overtime or the shoot-out) or zero points (if the loss occurs in regulation time).

As mentioned above, hockey teams tend to be built on lines. However, not all lines perform the same role during a game. A hockey team might have two scoring lines (players who are focused on scoring in offensive situations), one checking line (players who are focused on playing against the top scoring line of the opponent and are trying to limit their scoring opportunities), and one versatile line (in-between a scoring and checking line). Different game situations also determine the types of players sent out on the ice for a particular shift. In hockey, penalties committed by a player (e.g., tripping an opposing player with the hockey stick) result in the offending player going to the penalty box for at least two minutes or until the opposing team scores a goal, depending on the type and severity of the penalty. During this period, the offending team is on a "penalty kill" or is "short-handed" because it is one player short; the other team is on a "power play" because it has a one-player advantage. In these situations, the power-play team will typically start off with its best offensive players on the ice—sometimes the team will play four forwards and one defenseman. The short-handed team will counter with a mix of its best defensive players (both forwards and defensemen).

Hockey Statistics

Offensive players are typically characterized by their point production—either goals scored or assists on goals that teammates score. Defensive players should be judged by their ability to prevent goals. A composite statistic, plus-minus (+/-), can be used to measure both offensive and defensive contributions. Roughly speaking, this statistic measures a player's net contribution to goals scored both by his team and against

his team. Overall, the season-long performance statistics on which we focus for forwards and defensemen are

- Goals (G) = the number of times a player scores a goal;
- Assists (A) = the number of times a player assists on a goal; up to two players can assist on a goal and are counted as the last two players on the scoring team to touch the puck before the goal scorer;
- Points (Pts) = goals + assists;
- +/- = total goals scored by the player's team while he is on the ice (except in power-play situations) minus total goals scored by the opposing team while he is on the ice (except in short-handed situations);
- Hits = the number of hits a player delivers to opposing players;
- Blocks (Blks) = the number of opposing team shots a player blocks using part of his body or equipment;
- PIM = the number of penalty minutes a player is assessed for rule infractions;
- TOI = the amount of time the player is on the ice (also applies to goalies).

The statistics used to measure goalie performance differ significantly from those used for forwards and defensemen. We focus on the following:

- Save% = the number of shots the goalie stops divided by the total number of shots that he faces;
- Goals against average (GAA) = the number of goals scored on the goalie per 60 minutes of time on ice;
- Wins = the number of games in which the goalie is on the ice when his team scores the winning goal;
- Shutouts (SO) = the number of games in which the goalie does not allow a single goal in regulation time or in overtime (and therefore records a Save% of 1.0 and a GAA of 0.0);
- Games started (GS) = the number of games the goalie started (excludes games in which the goalie came on in relief of another goalie who started the game).

Data

The statistics we use in this paper are freely available to the public; we gathered them from <http://www.nhl.com>. We recorded data for all the statistics listed in

the previous section for NHL regular seasons from 2005–2006 to 2009–2010, inclusive. For example, in 2008–2009, we recorded statistics for 582 forwards, 303 defensemen, and 89 goalies across the 30 NHL teams.

Before doing our analysis, we culled our data set to eliminate players who played less than a certain threshold of minutes over the entire season. They were typically minor league players who were called up to play a limited number of NHL games or players who were injured and missed most of the season. We excluded these players to eliminate outliers whose performance over a short period would not have been indicative of season-long performance. The threshold for each position was determined by rank ordering the players of that position by TOI and eliminating the bottom 25 percent. In 2008–2009, the resulting thresholds were 170 minutes for forwards (~10 games), 326 minutes for defensemen (~16 games), and 410 minutes for goalies (~7 games).

Of the players who made the TOI cutoff each year, approximately 10 percent were traded over the course of that season. When determining player types via the clustering analysis, we used players' entire-season statistics. For traded players, we attributed a fraction of the player to each team for which he played, based on the number of minutes he played for each team. Although traded players could potentially have performed much better or worse after a trade, we believe such cases were in the minority and had a negligible impact on the final clustering results.

We obtained salary cap information for each player for the period 2007–2010 from <http://www.nhlnumbers.com>. A player's contribution to his team's salary cap is the same over all years of his contract. We obtained the annual amount (also referred to as the "cap hit") by dividing the total salary committed to the player over the life of his contract by the length of the contract in years.

Clustering Players into Distinct Types

To facilitate player comparisons on the same scale, we first normalized all forward and defensemen statistics using TOI. This accounted for some players playing more minutes per game than others. Next, we standardized each statistic by subtracting the average of

that statistic over all players of that position and then dividing by the associated standard deviation. Therefore, we were able to create a similar scale for all the statistics and prevent one from dominating the clustering process (Vincent and Eastman 2009). For goalie statistics, we normalized wins and shutouts by GS and then standardized all four statistics.

The clustering of players was conducted separately for each year. Statistics used in the clustering analysis were standardized season-long statistics. To create player clusters, we used the *k*-means clustering algorithm (Anderberg 1973) in MATLAB. We ran 1,000 replications of the algorithm for values of *k* from 2 to 10 (see Table 1 for the statistics used in the algorithm for each position).

To help determine the optimal number of clusters, we evaluated the Calinski-Harabasz pseudo-F index (Calinski and Harabasz 1974) and the silhouette value (Rousseeuw 1987) of different clusterings. Both values capture the tightness of the clusters—higher values indicate the presence of more distinct clusters. The pseudo-F statistic is essentially a ratio of the sum of squares of the distances between the cluster centroids and the mean of all the data points, and the sum of squares of the distances between data points and the centroid of the cluster to which a data point belongs. The silhouette value for a data point is the ratio of the difference between the minimum average distance to points in another cluster and the average distance to other points in the same cluster, and the maximum value of the two preceding values. The silhouette value of the overall data set is the average of the silhouette values for each data point. We used the F-statistic and silhouette values to guide us to an appropriate range for *k*. We found that $k \in \{2, 3, 4, 5\}$ had the highest values of these statistics, indicating

Forwards	Defensemen	Goalies
G/TOI	Pts/TOI	Save %
A/TOI	+/-/TOI	GAA
+/-/TOI	Hits/TOI	Wins/GS
Hits/TOI	Blks/TOI	SO/GS
Blks/TOI	PIM/TOI	
PIM/TOI		

Table 1: The table lists the statistics used in the clustering of players for each position.

	No. of players	G	A	+/-	Hits	Blks	PIM
Top Line	100	24.1 (1.05)	35.1 (1.20)	11.9 (1.11)	61.9 (-0.56)	24.9 (-0.53)	42.5 (-0.35)
Second Line	166	14.8 (0.12)	20.7 (0.12)	-4.2 (-0.20)	66.1 (-0.38)	23.7 (-0.38)	40.0 (-0.31)
Defensive	128	6.9 (-0.66)	9.3 (-0.74)	-4.3 (-0.38)	79.3 (0.29)	32.9 (0.98)	39.0 (-0.16)
Physical	43	3.4 (-0.94)	4.5 (-1.05)	-4.2 (-0.66)	103.5 (1.91)	11.3 (-0.21)	119.3 (2.48)

Table 2: The table illustrates four forward clusters and their defining characteristics. Values represent the mean value of the statistic in each cluster. The values in parentheses represent the mean value of the normalized (by TOI) and then standardized statistic. The best values in each column are highlighted, indicating dimensions that distinguish the clusters.

	No. of players	Pts	+/-	Hits	Blks	PIM
Offensive	70	36.4 (1.20)	3.3 (0.27)	63.6 (-0.64)	85.6 (-0.61)	48.4 (-0.32)
Defensive	71	14.6 (-0.49)	6.0 (0.62)	95.2 (0.23)	114.3 (0.76)	46.5 (-0.09)
Average	60	13.1 (-0.46)	-8.4 (-0.88)	62.7 (-0.25)	70.9 (-0.28)	34.2 (-0.32)
Physical	27	8.8 (-0.80)	-2.9 (-0.39)	116.9 (1.59)	73.6 (0.21)	79.6 (1.78)

Table 3: The table illustrates four defensemen clusters and their defining characteristics. Values represent the mean value of the statistic in each cluster. The values in parentheses represent the mean value of the normalized (by TOI) and then standardized statistic. The best values in each column are highlighted, indicating dimensions that distinguish the clusters.

that the most natural cluster sizes were within this range. At this point, we evaluated the set of clusters generated by each value of k —our choice for the optimal k for each position was guided by defining player types, which were separated along meaningful and natural dimensions from a hockey perspective. Tables 2, 3, and 4 show the clustering results for each position for the 2008–2009 season. Appendix A shows example players from each cluster.

For the forwards, we chose four clusters: Top Line F, Second Line F, Defensive F, and Physical F. Each category has a unique makeup and aligns with generally accepted player types. These categories are similar to those in Vincent and Eastman (2009); however, we have replaced their grinder category

with the more granular Second Line and Defensive categories. Notice that the forward clusters provide an almost perfect ordinal classification scheme with respect to the standardized statistics. G/TOI , A/TOI , and $+/-/TOI$ are decreasing when going from Top Line to Physical forwards, whereas $Hits/TOI$ and PIM/TOI are increasing. The $Blks/TOI$ statistic peaks in value at the Defensive cluster, as expected. The Second Line cluster is neither the best nor the worst in any one dimension.

For defensemen, we chose four clusters: Offensive D, Defensive D, Average D, and Physical D. Because each team normally has three pairs of defensemen, we initially expected three clusters to emerge. However, because the Offensive D and Physical D clusters were so well separated, choosing $k = 3$ would have resulted in a significant loss of granularity (Defensive D and Average D would have been essentially combined into one). Our results differ from those of Vincent and Eastman 2009, because their two defensemen clusters focused on scoring and physical play. Here, we see a clear defensive player type, as one would expect for defensemen.

Finally, for goalies, three clusters emerged: Elite G, Average G, and Bottom G. The goalie clusters have a different interpretation from the forward and

	No. of players	Save %	GAA	Wins/GS	SO/GS
Elite	18	0.92 (0.93)	2.40 (-1.01)	0.59 (0.93)	0.11 (1.30)
Average	35	0.91 (0.08)	2.79 (-0.09)	0.47 (-0.02)	0.04 (-0.31)
Bottom	14	0.89 (-1.39)	3.48 (1.52)	0.32 (-1.14)	0.01 (-0.89)

Table 4: The table illustrates three goalie clusters and their defining characteristics. Values represent the mean value of the statistic in each cluster. The values in parentheses represent the mean value of the standardized statistic. The best values in each column are highlighted, indicating dimensions that distinguish the clusters.

defensemen categories. Because only one goalie is on the ice at any time and his job is largely one dimensional (i.e., stop shots and prevent goals), our clusters simply separate good from average from poor performers; the statistics go from good to bad as goalies go from the Elite to Bottom cluster, thus supporting this observation.

We chose the same values of k for each position for each of the five seasons considered. The defining characteristic(s) for each cluster remained the same over all seasons, except in the case of Offensive D and Defensive D with respect to the $+/-$ /TOI statistic. The leader in $+/-$ /TOI among defensemen oscillated between Offensive D and Defensive D over the five years considered. However, the other defining statistic for those two clusters (Pts/TOI for Offensive D,

Blks/TOI for Defensive D) was consistent throughout all five seasons.

A Regression Model Relating Player Type to Team Performance

Having established the player clusters, we developed a regression model to determine a quantitative relationship between different player types and their effect on a team’s performance. First, we computed the average lineup on the ice for each team, in terms of the previously defined player types, for all seasons independently. For example, Table 5 lists the 30 NHL teams, and for each team shows the average number of players in each cluster who were on the ice at any time during the 2008–2009 regular season. The teams

Team	Pts.	Play-off	Forwards				Defense				Goalie		
			Top	Sec.	Def.	Phys.	Off.	Def.	Avg.	Phys.	Elite	Avg.	Bot.
San Jose	117	Y	1.5	0.7	0.5	0.1	1.4	0.4	0.0	0.1	1.0	0.0	0.0
Boston	116	Y	1.8	0.5	0.4	0.2	1.0	0.9	0.0	0.0	0.7	0.3	0.0
Detroit	112	Y	1.5	0.9	0.4	0.0	1.3	0.9	0.2	0.0	0.5	0.5	0.0
Washington	108	Y	1.1	1.1	0.6	0.1	0.4	1.2	0.0	0.2	0.0	0.9	0.0
New Jersey	106	Y	1.5	0.7	0.5	0.3	0.7	1.1	0.2	0.0	0.8	0.2	0.0
Chicago	104	Y	1.9	0.4	0.4	0.1	1.2	0.7	0.0	0.0	0.5	0.5	0.0
Vancouver	100	Y	1.2	1.0	0.6	0.1	0.9	0.7	0.0	0.2	0.6	0.3	0.0
Pittsburgh	99	Y	1.2	1.0	0.5	0.1	0.6	0.9	0.1	0.3	0.0	1.0	0.0
Philadelphia	99	Y	1.5	0.7	0.4	0.2	0.4	1.4	0.2	0.0	0.0	1.0	0.0
Calgary	98	Y	1.6	0.8	0.4	0.1	1.0	0.3	0.1	0.6	0.0	0.9	0.1
Carolina	97	Y	0.9	1.2	0.7	0.0	1.0	0.8	0.1	0.0	0.8	0.2	0.0
NY Rangers	95	Y	0.3	1.3	1.1	0.2	0.0	0.3	1.7	0.0	0.0	1.0	0.0
Montreal	93	Y	0.6	1.1	1.1	0.1	0.7	0.7	0.0	0.6	0.0	1.0	0.0
Florida	93	N	1.0	1.3	0.6	0.0	0.8	0.9	0.3	0.0	1.0	0.0	0.0
St. Louis	92	Y	0.7	1.3	0.8	0.1	0.5	0.3	1.1	0.0	0.7	0.0	0.3
Columbus	92	Y	0.6	1.3	0.8	0.2	0.4	1.3	0.2	0.0	0.7	0.0	0.1
Anaheim	91	Y	1.4	0.3	1.0	0.2	1.0	0.3	0.5	0.3	0.5	0.5	0.0
Buffalo	91	N	0.4	1.9	0.4	0.1	0.8	0.5	0.5	0.0	0.7	0.0	0.3
Minnesota	89	N	0.3	1.7	0.6	0.3	0.9	0.3	0.7	0.1	0.8	0.2	0.0
Nashville	88	N	0.8	0.9	1.0	0.2	0.9	0.3	0.8	0.0	0.6	0.4	0.0
Edmonton	85	N	0.3	1.9	0.5	0.1	1.3	0.3	0.0	0.4	0.0	1.0	0.0
Ottawa	83	N	0.3	1.4	1.0	0.1	0.5	0.5	0.8	0.1	0.0	1.0	0.0
Dallas	83	N	0.7	1.6	0.5	0.1	0.7	0.7	0.4	0.1	0.0	0.9	0.1
Toronto	81	N	0.6	1.5	0.7	0.1	0.9	0.6	0.0	0.4	0.0	0.1	0.8
Phoenix	79	N	0.4	1.7	0.6	0.2	0.2	0.9	0.8	0.0	0.0	0.9	0.0
Los Angeles	79	N	0.0	2.0	0.5	0.3	0.3	0.0	1.2	0.5	0.5	0.5	0.0
Atlanta	76	N	1.1	1.0	0.7	0.1	1.1	0.3	0.0	0.5	0.0	0.5	0.5
Colorado	69	N	0.0	2.0	0.7	0.3	0.3	0.6	0.9	0.0	0.0	0.7	0.3
Tampa Bay	66	N	0.4	1.4	0.9	0.2	0.1	0.1	1.1	0.3	0.0	0.5	0.5
NY Islanders	61	N	0.0	1.4	1.4	0.0	0.4	0.1	1.1	0.3	0.0	0.4	0.6

Table 5: The table shows 2008–2009 NHL teams’ regular season performance and mix of player types. Numerical values represent the average number of players in each player category over the course of the season.

are ordered based on the number of points obtained at the end of the season.

To calculate these numbers, we first determined the cluster to which each player belonged and divided his total time on ice played over the whole season by $82 \times 60 + 47 = 4,967$ (82 games per season, each with 60 minutes of regulation time, plus an approximate number of overtime minutes in a season; therefore, an average game lasts 60.6 minutes). Then, we added up this fraction over all players of the same type on the same team. For example, a Top Line forward from San Jose who played 1,500 minutes in the season would contribute $1,500/4,967 = 0.30$ to that cell. Intuitively, the numbers for a particular team represent the average lineup the team had on the ice over the year. We calculated the approximate number of overtime minutes per season based on the average number of overtime games played per team over 2005–2010, multiplied by 5 minutes. In general, this will be a slight overestimation, because many overtime games end before the 5 minutes is up. However, overtime minutes account for only about 1 percent of the total minutes played in a season. Using a sensitivity analysis, we found that our results were insensitive to changes in the estimate of overtime minutes.

Next, we conducted a multiple linear regression using the player types as independent variables and the number of team points obtained in the regular season as the dependent variable. We used the 2009–2010 data as the validation year and trained the regression model using the data from the previous four seasons (2005–2006 to 2008–2009). We tested two models: one that included only variables with significant coefficients (regression 1) and one with all 11 variables (regression 2). In both cases, we did not include an intercept parameter, because we wanted as many of the player types as possible to be statistically significant; in supplemental tests, we found that models with an intercept variable had many fewer player types with significant coefficient values. Next, we compared the in-sample and out-of-sample performance for the two regression models. We measured performance by computing the root mean squared error (RMSE) between the predicted and actual team points, as well as calculating the percentage of cases in which the actual team points value fell within

the corresponding prediction interval. Table B.1 in Appendix B summarizes the results.

Regression 1 performed better than regression 2 in terms of out-of-sample RMSE (10.2 versus 10.8), as well as the degradation in RMSE when taking the ratio of out-of-sample over in-sample RMSE (1.5 versus 1.6). Regression 2 performed better in terms of the prediction interval metric. However, given the large number of insignificant variables in regression #2 and the better performance in RMSE of regression 1, we decided to use only the variables in regression 1 going forward. In a supplemental analysis, we also chose different years to be used as the validation year, leaving the remaining four years of data to train the model. We tried all $\binom{5}{1} = 5$ combinations, kept only significant variables, and found that the regression coefficient values were largely consistent across the years.

With the set of independent variables fixed to the ones used in regression 1, we reran the regression with all five years of data (2005–2010) to generate a final set of regression coefficients. We denote this final model as regression 3. Table 6 summarizes the results of regression 3, including confidence intervals for each significant coefficient. A regression coefficient can be thought of as the contribution of one player of the corresponding type to the total number of team points the team earns over the season. Specifically, the regression coefficient for a particular player type is the contribution to team points if a player of that type played 60.6 minutes a game for all 82 games in the season. For example, a Top Line F coefficient of 28.8 means that a player classified as a Top Line F, if he averages 20.2 minutes a game (a third of a game) and plays all season, will contribute about 9.6 team points to the total number of points his team earns over the course of the season. In general, the contribution to team points of a player of type i who played x minutes over the regular season is calculated as $(x/4,967)\beta_i$, where β_i is the regression coefficient corresponding to player type i . Table 6 also includes salary cap information. The reported salary cap numbers are the average salary cap hit values of all players in each cluster from 2007 to 2010.

Four player types—Physical F, Average D, Physical D, and Bottom G—were statistically insignificant in regression 3 (and in regression 1), which indicates that

Position	Cluster	Team points	Cap hit (\$M)	Team pts/ Cap hit	ATOI (min)	Team pts/Cap hit ×ATOI/60.6
Forward	Top Line	28.8 (26.0, 31.6)	3.36 (2.12)	8.6 (7.7, 9.4)	17.73	2.5 (2.3, 2.8)
	Second Line	19.9 (17.9, 21.9)	2.19 (1.61)	9.1 (8.2, 10.0)	15.32	2.3 (2.1, 2.5)
	Defensive	18.7 (15.3, 22.0)	1.10 (0.86)	17.0 (13.9, 20.0)	13.07	3.7 (3.0, 4.3)
	Physical	—	0.73 (0.31)	—	8.15	—
Defense	Offensive	9.3 (5.7, 12.9)	3.17 (2.15)	2.9 (1.8, 4.1)	21.74	1.1 (0.6, 1.5)
	Defensive	3.7 (0.0, 7.5)	1.99 (1.16)	1.9 (0.0, 3.8)	19.36	0.6 (0.0, 1.2)
	Average	—	1.90 (1.45)	—	19.73	—
	Physical	—	1.48 (1.17)	—	16.94	—
Goalie	Elite	32.1 (27.4, 36.8)	2.72 (2.23)	11.8 (10.0, 13.5)	58.10	11.3 (9.7, 13.0)
	Average	21.2 (16.2, 26.2)	2.18 (1.86)	9.7 (7.4, 12.0)	56.81	9.1 (7.0, 11.3)
	Bottom	—	1.79 (1.40)	—	54.77	—

Table 6: The table illustrates the average salary cap hit and contribution to team points for each player type. Values in the team points column correspond to the regression coefficients. The team points values in parentheses are the 95% confidence intervals for the regression coefficient (note that the Defensive D variable had a p -value of 0.05). We used “—” to identify player types that were not included in the regression. The values in parentheses in the “Cap hit” column represent the standard deviation of cap hits in that cluster. The values in the ATOI column represent the average time on ice in minutes that players of each player type received.

these player types do not make a significant contribution to the number of team points their team earns over the course of the regular season. Adding any of these four types back into the regression would result in at least one of the independent variables losing statistical significance. We believe that the remaining seven player types impact team performance the most. Although we do not have statistically significant regression coefficients for 4 of the 11 player types, less than 20 percent of all players are in these clusters and their combined ice time is less than 15 percent of the total ice time played by all players.

The regression results highlight a few interesting observations. First, Elite goalies seem to contribute the most to team points, which is somewhat intuitive. Only one goalie per team is on the ice at any moment; he is the last line of defense in preventing opposition goals. As expected, Top Line forwards contribute more to team points than Second Line forwards, who in turn contribute more than Defensive forwards. We can see similar trends among the defensemen clusters and goalie clusters. By evaluating the range of the 95 percent confidence intervals for each regression coefficient, we can also observe more relative variability in the contributions of defensemen than in that of forwards or goalies. For example, the lower end of the 95 percent confidence interval for Top Line forwards (26.0) is 90 percent of the coefficient (28.8), whereas the lower end of the 95 percent confidence interval for

Offensive defensemen (5.7) is 61 percent of the coefficient (9.3). This relative variability is even larger for the Defensive D cluster. These observations suggest that finding high-performing defensemen might merit more focus on the part of GMs than finding high-performing forwards, especially because a team normally has only half as many defensemen as forwards.

Are Certain Player Types Overvalued or Undervalued?

The economic burden that each player contributes to his team’s salary cap is an issue that we cannot ignore in the NHL today. Using the results of the regression analysis, available salary cap information, and playing-time information, we computed a value measure for each player type. First, we divided the average contribution to team points, as well as the corresponding confidence interval endpoints, by the average salary cap hit of players in that cluster (see column Team Pts/Cap Hit in Table 6). Then, we multiplied the result by the average fraction of ice time per game (out of 60.6 minutes) given to a player of that type (see last column in Table 6). For each cluster, we calculated this fraction as the average time on ice (ATOI) in minutes for players in the cluster divided by 60.6. We calculated the ATOI for each cluster as the average ice time per game of all players in that cluster over all five seasons from 2005–2010. It is interesting to observe that of the forward types,

Defensive forwards provide the most value for their contributions. Despite being inferior to Top Line and Second Line forwards in terms of contribution to team points, Defensive forwards appear to be a good investment primarily because of their low cost. However, the value of Top Line forwards, who are typically the best players on a hockey team, seems to be dragged down by their high salaries. If we consider only cap hits, forwards would seem to provide approximately the same amount of value as goalies. However, because the regression coefficients represent the contribution to team points if a player plays every minute of every game, it becomes evident that goalies provide the most value (on a per-game basis) of all the player types. These results also highlight that when we consider salaries and differences in playing time, certain player types with different contributions to team points (e.g., Top Line and Second Line forwards) are more or less on an equal footing. A similar relationship exists between Elite and Average goalies.

On average, defensemen seem to be overpaid in comparison to other player types, especially when we consider the variability in their performance. However, given the observed higher variability in the defense clusters as compared to the forward clusters, GMs should be willing to pay more for established and proven defensemen. The relatively lower variability in the Top Line F cluster suggests that overpaying for talent in this cluster may not make sense from strictly a team performance perspective. However, in practice, even if Top Line forwards do not provide the best marginal value based on salary and playing time, teams with available salary cap space would still likely prefer to have as many of them as possible to maximize their competitiveness.

A Trade Tool

Based on the results of the regression model, we developed an Excel-based trade tool to evaluate the impact of a potential trade based on the players involved, the existing makeup of the team, and the projected ice-time distribution for the team's players. The trade tool is available as an electronic companion to this paper, which is part of the online version that can be found at <http://interfaces.journal.informs.org/>. We believe this trade tool can

be a decision support tool for a GM or coach. The user can evaluate the contributions to team performance (in terms of team points) of different rosters by adjusting the players on the team and, most importantly, the ice time given to those players. Determining the amount of ice time players receive is normally the decision of the coaching staff; therefore, it is reasonable to expect that these decision makers would need to estimate changes in ice-time allocations as a result of roster changes. For the purposes of evaluating a trade, the coach and GM should discuss the potential ramifications of changes in playing time for the players affected.

We emphasize the need for trade evaluations to account for the existing makeup of the team and how the ice time is redistributed after the trade. For example, trading away a Top Line forward (coefficient of 28.8) for an Average goalie (coefficient of 21.2) might not seem like a good idea when framed as a simple swap. However, if the Average goalie were to take over from a Bottom goalie for the remainder of the season, and if other Top Line or Second Line forwards would get more ice time because of the departure of the traded forward, this trade may actually be balanced or even beneficial to the team. As another example, a team may be willing to trade away an Average goalie if it has a backup Average goalie who can eat up the minutes of the departed goalie, and if the acquired player (e.g., a Top Line forward) ends up playing minutes that would otherwise be given to a lower-tier forward.

Relating Trades to Changes in Team Performance

We used the trade tool to evaluate trades involving the Phoenix Coyotes in both the 2008–2009 and 2009–2010 seasons. We chose to evaluate the Phoenix Coyotes because all its trades in those two years were concentrated on the trade deadline day. This provided a relatively clean separation between the team's pre- and post-trade composition. Of course, our tool can be used by teams with multiple trades throughout the year; all that one needs to do is to put more effort into gathering and parsing the data for all of the pre- and post-trade compositions for each trade considered. In this subsection, our aim is more about validating the tool; therefore, we used actual ice-time statistics both before and after the trade to evaluate whether

the change in team composition can be related to the change in team performance.

In our first example, we examine the trades Phoenix made at the 2008–2009 trade deadline. In separate trades with four different teams (Calgary, Philadelphia, New York Rangers, and Buffalo) at the trade deadline, Phoenix acquired Matthew Lombardi (Top Line F), Brandon Prust (Physical F), Scottie Upshall (Second Line F), Dmitri Kalinin (Average D), Nigel Dawes (Defensive F), and Petr Prucha (Defensive F) for Olli Jokinen (Second Line F), Derek Morris (Average D), Daniel Carcillo (Physical F), and Mikael Tellqvist (Average G). As a result of these trades, Phoenix added one Top Line forward and two Defensive forwards in exchange for an Average goalie. Prior to the trades, Phoenix had accumulated 59 team points in 63 games, which put the team on pace for 76.8 points for the season. Phoenix finished the season with 79 points in total—a 2.2 point increase over the pace estimate. Based on Phoenix's pre-trade versus post-trade team composition and the known ice-time allocations, our model suggests that Phoenix would essentially break even in team performance (a gain of 0.1 team points) as a result of the trades. The addition of the forwards benefited Phoenix because they were able to play minutes that would otherwise be given to lower-tier players. However, trading away Mikael Tellqvist (Average G) and having Al Montoya (an unclustered goalie because of minimal playing time), who is not shown in the tool, play in 5 of the remaining 19 games resulted in a projected decrease in team performance. We note that Al Montoya was classified as an Average G in the following season. At the end of the 2008–2009 season, he was being groomed to be the next backup goalie for Phoenix. If we include him in this trade evaluation as an Average G for the 5 games he played, then our model predicts that Phoenix would have improved by an extra 1.2 team points following the trade. In terms of salary, these trades resulted in a reduction of Phoenix's cap hit by almost \$4.6 million (annualized). Overall, although the trade might have only had a marginal impact on the Phoenix team's performance, Phoenix did benefit significantly in terms of a reduction in salary obligations. Appendix C shows a summary of this trade.

The following year, in 2009–2010, Phoenix again made a number of moves at the trade deadline. In a

series of trades with six teams, Phoenix netted a Top Line F, an Offensive D, and an Average D. At that point in the season, Phoenix had played 64 games and had 79 team points in the standings, which put the team on pace for 101.2 points at the end of the season. Phoenix finished with 107 team points—an improvement of 5.8 points over the pace estimate. Our tool suggests that Phoenix received a net benefit of 1.7 team points because of the trade (predicted team points accumulated in the final 18 games accounting for the trade minus the predicted team points accumulated in the final 18 games without the trade) and would finish the season with 101.9 points. Although this is lower than the actual improvement based on the original pace estimate, the model correctly identified a net team point increase based on the post-trade team composition. Furthermore, the difference between the predicted and actual team points value was $107 - 101.9 = 5.1$, which was less than the standard error (7.0) of the regression.

Comparing Trade Alternatives

In this subsection, we evaluate a well-known trade between the Ottawa Senators and the San Jose Sharks. In the summer of 2009, Dany Heatley requested a trade out of Ottawa. The situation was complicated because Heatley had a no-trade clause in his contract, allowing him to block any trade he did not like. Edmonton was the first team to table an offer (in June); however, Heatley rejected it. It was not until September that Ottawa finalized a trade with San Jose, which Heatley accepted. Many people believed that the Edmonton deal was better, and that Ottawa was forced to make a suboptimal deal in an effort to avoid roster distractions entering its 2009–2010 training camp. We demonstrate how our trade tool can be used to estimate which trade would have been more beneficial to Ottawa.

Using the 2008–2009 clusters, Heatley was classified as a Second Line F (but was a Top Line F in all previous years). Edmonton would have sent Dustin Penner (Second Line F), Andrew Cogliano (Second Line F), and Ladislav Smid (Physical D) to Ottawa for Heatley and a second-round draft pick in the 2010 entry draft. San Jose offered Milan Michalek (Top Line F), Jonathan Cheechoo (Second Line F), and a second-round draft pick for Heatley and a

fifth-round draft pick. The timing of these trades also had secondary effects on Ottawa's roster. If the Edmonton trade had been made in June, analysts and reporters believed that Ottawa would have then signed free agent Mike Cammalleri (Top Line F) at the start of the free agency period on July 1. However, when the Edmonton trade fell through, the uncertainty surrounding Ottawa's financial situation made Cammalleri too expensive for Ottawa. Cammalleri signed with Montreal on July 1; Ottawa signed Alexei Kovalev (Second Line F), also a free agent, five days later. Using our trade tool, we analyzed both the Edmonton and San Jose trade offers (including the secondary effect of Cammalleri versus Kovalev) using the 2009–2010 Ottawa roster. When using the 2009–2010 ice-time allocations for all players, we found that the Edmonton deal would have netted Ottawa 1.3 team points over the San Jose deal. The improvement is primarily because of the increased ice time the Edmonton players got over the San Jose players. The ice-time differential is also reflective of the injury histories of Michalek and Cheechoo in comparison to those of Cogliano and Penner; Michalek and Cheechoo have more history of missing games because of injury. When we used the 2008–2009 ice-time allocations, we found that the Edmonton deal would have netted Ottawa 0.4 team points over the San Jose deal. Furthermore, if we use the players' 2009–2010 clusters (with the 2009–2010 ice-time allocations), the Edmonton deal would have netted Ottawa 6.3 team points over the San Jose deal. This increase also reflects that the San Jose players involved in the deal performed much worse in 2009–2010 than in 2008–2009. Overall, across these different playing-time and clustering scenarios, it appears that the Edmonton deal would have benefitted Ottawa more than the San Jose deal did.

Discussion and Extensions

Clustering is a simple way to summarize a player's type based on his performance along multiple statistical dimensions. Often, looking at a single statistic (e.g., Pts/TOI) can be misleading. For example, Logan Couture played 20 games in 2009–2010 and was classified as a Top Line F; however, he was ranked 137th in Pts/TOI (below 44 Second Line for-

wards). In 79 games in 2010–2011, he was second in points and goals among rookies; he was also nominated for the rookie of the year award. A decision maker who only considers 2009–2010 Pts/TOI might overlook or misclassify Couture. However, a player who has a high Pts/TOI might not necessarily be clustered as a Top Line F. For example, Vincent Lecavalier was clustered as a Second Line F in 2009–2010, but had a Pts/TOI statistic that was higher than 46 (out of 95) Top Line forwards. Clustering provides a more comprehensive picture of a player in comparison to using a single metric.

Classifying players into distinct types can be viewed as a strategic analysis to understand the mix of player types on NHL teams. Of course, variability within the clusters suggests that additional insight might be gained by isolating and quantifying the value of individual players. Using our clustering approach, it might be possible to generate a composite picture of each player based on distances to different clusters—a split-personality classification, so to speak. A tool might be developed to focus even more on the individual characteristics of the players involved and provide a player-specific valuation. This is an area for future study.

Although we demonstrate the use of the trade tool on historical trades in this paper, we believe this tool has more value in evaluating prospective trades. Admittedly, projecting exactly how much ice time players will get, how many games each player will play, who will get injured, or what player performance will be like in the future is impossible. However, this tool allows a user to test different clustering and ice-time scenarios and estimate the corresponding impact on the team's performance. For example, in the Heatley trade, we used 2009–2010 TOI to evaluate the trades to simulate the act of a GM projecting future playing time.

Age and past performance can be important trade considerations that our clustering and regression analysis do not currently capture. For example, consider a Pittsburgh-Anaheim trade made days prior to the 2008–2009 trade deadline. Pittsburgh received Chris Kunitz (Top Line F) and Eric Tangradi (prospect) in exchange for Ryan Whitney (Average D).

The trade filled a gap for both teams; however, Pittsburgh seemed to benefit more, both in terms of contribution to team points and salary cap (Kunitz's cap hit was \$3.725 million for three more years; Whitney's cap hit was \$4 million for four more years). Incidentally, Kunitz played a big role in Pittsburgh's Stanley Cup championship that year, because he tied for fourth in team scoring during the play-offs. Based on these players' clusters alone, it would appear that Anaheim overpaid for Whitney and perhaps could have asked for more in return for Kunitz (e.g., avoid giving up Tangradi). However, Anaheim might have accepted this trade because Whitney was four years younger than Kunitz and would have had time to regain his previous form (Whitney was classified as an Offensive D in years prior to 2008–2009). This trade also reinforces the idea that managers might be willing to pay more for a defenseman who has consistently shown strong performance in the past.

Regarding the available salary cap information, because the league-wide salary cap has been increasing every year since the lockout, the cap numbers reported in Table 6 are most likely overestimates of the true average from 2005–2010 (the range of data used to generate the regression coefficients). The average growth rate in cap hit over 2007–2010 ranged from –1 percent (Defensive D) to +16 percent (Top Line F), which suggests that the value ratios would be slightly higher had we used the average salary cap over 2005–2010 instead of 2007–2010. However, even with the 2005–2010 cap data, the relative ordering of the numbers across the clusters should be similar to our computations.

The financial success of a professional sports organization is often linked to wins and losses. Winning teams may have higher attendance at their games and their fans may spend more; they are also more likely to make the playoffs, generating additional revenue from ticket and concession sales. The marginal monetary value of a win can be thought of as the additional revenue that is gained by a team for each win it achieves (Schmidt and Berri 2001, Burger and Walters 2003, Gennaro 2007, Fleisher et al. 1992). Using this concept, we can relate the value of a player to a potential increase in team earnings through his contribution to team points. Of course, a player's economic

contributions to his team go beyond wins and losses. Although the salaries of Top Line forwards might seem high if we consider only their contributions to team points, they likely provide additional economic value to the organization (e.g., through merchandise sales and promotional activities). The superstar effect, as Lucifora and Simmons (2003) describe in soccer, might simply be the price of trying to gain a competitive advantage both on the ice and in the financial statements.

Conclusions

In this paper, we use *k*-means clustering to create a novel classification scheme for NHL hockey players based on a variety of statistics, and we develop a regression model to quantify the effect of different player types on team performance. We identify four forward types, four defensemen types, and three goalie types. Our regression indicates that goalies contribute the most to team performance, followed by forwards and then defensemen. We also find that when we consider player salaries and ice-time allocations, player types previously considered less attractive (e.g., defensive-oriented forwards) become investments with a comparable return (because of their low salaries) to more offensive-oriented forwards. Finally, we package the results of our regression model into an Excel-based tool that can be used to estimate the impact that trades or other roster changes can have on a team's performance.

Sports analytics is a burgeoning field and we believe that hockey is a source of many fascinating problems to which we can apply sophisticated analytical methods. We encourage decision makers in the hockey industry to deepen their use of analytics and to use it to explore new opportunities to create a competitive advantage for their teams.

Appendix A. Examples of Players in Each Cluster

Table A.1 lists six example players from each player cluster. To provide more granularity, we classified players as high (i.e., cap hit values in the top third in their cluster), medium (i.e., cap hit values in the middle third in their cluster), or low (i.e., cap hit values in the bottom third in their cluster) with respect to their salary cap hit in 2008–2009.

Position	Player type	Cap hit		
		High	Medium	Low
Forward	Top Line	Sidney Crosby Alexander Ovechkin	Zach Parise Henrik Zetterberg	Loui Eriksson Bobby Ryan
	Second Line	Scott Gomez Brad Richards	Ryan Kesler Wojtek Wolski	Brandon Dubinsky James Neal
	Defensive	Kris Draper Chris Drury	Maxim Lapierre Maxime Talbot	Colin Fraser Darren Helm
	Physical	Daniel Carcillo Chris Neil	Cal Clutterbuck Brad May	George Parros Brandon Prust
Defense	Offensive	Mike Green Nicklas Lidstrom	Paul Martin Mark Streit	Duncan Keith Kris Letang
	Defensive	Robyn Regehr Brent Seabrook	Hal Gill Anton Volchenkov	Andy Greene Rob Scuderi
	Average	Ed Jovanovski Wade Redden	Francois Beauchemin Jordan Leopold	Chris Chelios Marc Staal
	Physical	Mike Komisarek Brooks Orpik	Shane O'Brien Ladislav Smid	Sheldon Brookbank John Erskine
Goalie	Elite	Martin Brodeur Roberto Luongo	Ryan Miller Tim Thomas	Craig Anderson Pekka Rinne
	Average	Ilya Bryzgalov Marty Turco	Antero Niittymaki Carey Price	Jaroslav Halak Michael Leighton
	Bottom	Pascal Leclaire Vesa Toskala	Patrick Lalime Andrew Raycroft	Joey MacDonald Curtis McElhinney

Table A.1: The table shows player examples from each cluster for 2008–2009.

Appendix B. Regression Models

Table B.1 summarizes the regression models we described in the section *A Regression Model Relating Player Type to Team Performance*.

Regression no.	1	2	3
Top Line F	27.6	31.4	28.8
Second Line F	18.5	21.6	19.9
Defensive F	18.8	22.5	18.7
Physical F	—	18.5	—
Offensive D	8.9	13.6	9.3
Defensive D	6.7	11.1	3.7
Average D	—	4.1	—
Physical D	—	6.8	—
Elite G	33.7	12.2	32.1
Average G	22.0	0.9	21.2
Bottom G	—	−19.4	—
In-sample RMSE	6.7	6.6	7.0
Out-of-sample RMSE	10.2	10.8	N/A
In-sample P.I. (%)	94.2	96.7	94.7
Out-of-sample P.I.	86.7%	93.3%	N/A

Table B.1: In our regression models, highlighted cells indicate variables that we included in the regression model but were not significant at the 95 percent level (i.e., had a p -value > 0.05). The prediction interval (P.I.) percent represents the fraction of team point values that fell within their corresponding prediction intervals.

Appendix C. Trade Tool

In Figures C.1 and C.2, we show two images that summarize the output of trades we evaluated using our Excel-based trade tool.

Note that in Figure C.1, Kevin Porter (Second Line F) was sent to Phoenix’s minor league affiliate earlier in the season prior to the trade deadline. In both figures, differences in the roster between the left and right sides of the table are highlighted in grey. For example, in Figure C.2, Brian Lee is highlighted on the left side because Ottawa’s roster had room for him after the San Jose trade; on the right side, we assumed that as a result of the acquisition of Ladislav Smid, Brian Lee would have remained with Ottawa’s minor league affiliate. The average ice times of the Defensive and Physical forwards were reduced in the Edmonton side of the Heatley trade (compared to the San Jose side) to account for the additional minutes and games that the acquired forwards (Cammalleri, Cogliano, and Penner) would have played. Note that for goalies, the numbers in the GP column actually refer to games started (GS), and the ATOI calculation is based on total time on ice divided by GS.

Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://interfaces.journal.informs.org/>.

Pre-Trade								Post-Trade							
Player	Cap Hit (M)	Cluster	Cluster Coefficient	GP	ATOI	Accumulated Team Points	Projected Season Team Points	Player	Cap Hit (M)	Cluster	Cluster Coefficient	GP	ATOI	Accumulated Team Points	Projected Season Team Points
Forwards								Forwards							
Shane Doan	4.55	Top Line	28.8	63	20.87	7.62	9.92	Shane Doan	4.55	Top Line	28.8	19	18.11	2.00	8.61
Mikkel Boedker	1.73	Second Line	19.9	62	15.90	3.95	5.14	Matthew Lombardi	1.82	Top Line	28.8	19	20.90	2.30	9.94
Olli Jokinen	5.25	Second Line	19.9	57	18.16	4.15	5.40	Mikkel Boedker	1.73	Second Line	19.9	16	14.10	0.90	3.90
Joakim Lindstrom	0.50	Second Line	19.9	35	15.63	2.19	2.85	Joakim Lindstrom	0.50	Second Line	19.9	9	11.83	0.43	1.84
Enver Lisin	0.91	Second Line	19.9	43	15.04	2.59	3.37	Enver Lisin	0.91	Second Line	19.9	5	12.90	0.26	1.11
Peter Mueller	1.70	Second Line	19.9	55	16.83	3.71	4.83	Peter Mueller	1.70	Second Line	19.9	17	13.61	0.93	4.00
Kevin Porter	0.95	Second Line	19.9	34	13.62	1.85	2.41	Steven Reinprecht	2.00	Second Line	19.9	15	14.00	0.84	3.63
Steven Reinprecht	2.00	Second Line	19.9	58	16.39	3.81	4.96	Viktor Tikhonov	1.18	Second Line	19.9	15	12.72	0.76	3.30
Viktor Tikhonov	1.18	Second Line	19.9	46	11.92	2.20	2.86	Kyle Turris	2.70	Second Line	19.9	13	14.47	0.75	3.25
Kyle Turris	2.70	Second Line	19.9	50	12.51	2.51	3.26	Scottie Upshall	1.23	Second Line	19.9	19	18.59	1.42	6.11
Todd Fedoruk	1.06	Defensive	18.7	63	10.73	2.54	3.31	Nigel Dawes	0.59	Defensive	18.7	12	14.15	0.64	2.76
Martin Hanzal	0.98	Defensive	18.7	60	16.84	3.80	4.95	Todd Fedoruk	1.06	Defensive	18.7	9	9.59	0.32	1.40
Daniel Winnik	0.53	Defensive	18.7	35	13.18	1.74	2.26	Martin Hanzal	0.98	Defensive	18.7	14	14.24	0.75	3.24
Daniel Carcillo	0.89	Physical	0.0	54	11.98	0.00	0.00	Petr Prucha	1.60	Defensive	18.7	19	18.44	1.32	5.69
Forwards Total	24.92	14	715	10,836.90	42.66	55.53		Daniel Winnik	0.53	Defensive	18.7	14	12.79	0.67	2.91
Forwards Max			756	11,340.00				Brandon Prust	0.50	Physical	0.0	11	9.98	0.00	0.00
Defense								Defense							
Keith Yandle	0.85	Offensive	9.3	51	16.30	1.56	2.03	Keith Yandle	0.85	Offensive	9.3	18	17.50	0.59	2.55
Ken Klee	1.25	Defensive	3.7	49	14.50	0.53	0.69	Ken Klee	0.63	Defensive	3.7	19	17.60	0.25	1.08
Zbynek Michalek	1.25	Defensive	3.7	63	23.05	1.08	1.41	Zbynek Michalek	1.25	Defensive	3.7	19	21.56	0.31	1.32
Kurt Sauer	1.75	Defensive	3.7	49	20.69	0.76	0.98	Kurt Sauer	1.75	Defensive	3.7	19	20.43	0.29	1.25
David Hale	0.70	Average	0.0	43	15.04	0.00	0.00	David Hale	0.70	Average	0.0	5	15.94	0.00	0.00
Ed Jovanovski	6.50	Average	0.0	63	22.65	0.00	0.00	Ed Jovanovski	6.50	Average	0.0	19	20.50	0.00	0.00
Derek Morris	3.92	Average	0.0	57	21.26	0.00	0.00	Dmitri Kalinin	2.10	Average	0.0	15	20.20	0.00	0.00
Defense Total	16.22	7	375	7,292.97	3.92	5.11		Defense Total	13.78	7	114	2,219.67	1.43	6.19	
Defense Max			378	7,560.00				Defense Max			114	2,280.00			
Goalies								Goalies							
Ilya Bryzgalov	4.25	Average	21.2	50	59.75	12.75	16.60	Ilya Bryzgalov	4.25	Average	21.2	13	59.41	3.30	14.23
Mikael Tellqvist	0.80	Average	21.2	15	53.20	3.41	4.43	Goalies Total	4.25	1	13	772.33	3.30	14.23	
Goalies Total	5.05	2	65	3,785.47	16.16	21.03		Goalies Max			19	1,140.00			
Goalies Max			63	3,780.00				Overall							
Team Total	46.19	23	1,155	21,915.34	62.74	81.66		Team Total	41.58	24	353	6,398.13	19.03	82.12	
Team Max			1,197	22,680.00				Team Max			361	6,840.00			
Trade Impact								Trade Impact							
								-4.61							
								0.10							
								0.45							

Figure C.1: The screenshot shows Phoenix's pre- and post-trade team compositions in 2008-2009.

San Jose trade								Edmonton trade							
Player	Cap Hit (M)	Cluster	Cluster Coefficient	GP	ATOI	Accumulated Team Points	Projected Season Team Points	Player	Cap Hit (M)	Cluster	Cluster Coefficient	GP	ATOI	Accumulated Team Points	Projected Season Team Points
Forwards								Forwards							
Daniel Alfredsson	4.88	Top Line	28.8	70	19.65	7.98	7.98	Daniel Alfredsson	4.88	Top Line	28.8	70	19.65	7.98	7.98
Milan Michalek	4.33	Top Line	28.8	66	18.25	6.98	6.98	Michael Cammalleri	6.00	Top Line	28.8	81	17.55	8.24	8.24
Matt Cullen	2.88	Second Line	19.9	21	18.00	1.51	1.51	Matt Cullen	2.88	Second Line	19.9	21	18.00	1.51	1.51
Nick Foligno	0.85	Second Line	19.9	61	14.30	3.49	3.49	Nick Foligno	0.85	Second Line	19.9	61	14.30	3.49	3.49
Peter Folin	0.61	Second Line	19.9	75	12.88	3.87	3.87	Peter Folin	0.61	Second Line	19.9	75	12.88	3.87	3.87
Ryan Shannon	0.60	Second Line	19.9	66	12.67	3.35	3.35	Ryan Shannon	0.60	Second Line	19.9	66	12.67	3.35	3.35
Jason Spezza	7.00	Second Line	19.9	60	19.05	4.58	4.58	Jason Spezza	7.00	Second Line	19.9	60	19.05	4.58	4.58
Jonathan Cheechoo	3.00	Second Line	19.9	61	11.93	2.92	2.92	Andrew Cogliano	1.13	Second Line	19.9	82	18.06	5.93	5.93
Alex Kovalev	5.00	Second Line	19.9	77	18.15	5.60	5.60	Dustin Penner	4.25	Second Line	19.9	78	19.23	6.01	6.01
Shean Donovan	0.63	Defensive	18.7	30	7.42	0.84	0.84	Shean Donovan	0.63	Defensive	18.7	30	7.42	0.84	0.84
Mike Fisher	4.20	Defensive	18.7	79	18.97	5.64	5.64	Mike Fisher	4.20	Defensive	18.7	79	17.72	5.27	5.27
Chris Kelly	2.13	Defensive	18.7	81	14.95	4.56	4.56	Chris Kelly	2.13	Defensive	18.7	81	13.58	4.14	4.14
Jarkko Ruutu	1.30	Defensive	18.7	82	13.35	4.12	4.12	Jarkko Ruutu	1.30	Defensive	18.7	82	9.15	2.82	2.82
Jesse Winchester	0.55	Defensive	18.7	52	10.02	1.96	1.96	Jesse Winchester	0.55	Defensive	18.7	15	11.13	0.63	0.63
Chris Neil	2.00	Physical	0.0	68	11.97	0.00	0.00	Chris Neil	2.00	Physical	0.0	68	9.56	0.00	0.00
Forwards Total	39.94	15	949	14,262.15	57.40	57.40		Forwards Total	38.99	15	949	14,263.10	58.67	58.67	
Forwards Max			984	14,760.00				Forwards Max			984	14,760.00			
Defense								Defense							
Erik Karlsson	1.30	Offensive	9.3	60	20.10	2.26	2.26	Erik Karlsson	1.30	Offensive	9.3	60	20.10	2.26	2.26
Filip Kuba	3.70	Offensive	9.3	53	22.85	2.27	2.27	Filip Kuba	3.70	Offensive	9.3	53	22.85	2.27	2.27
Andy Sutton	3.00	Defensive	3.7	18	19.22	0.26	0.26	Andy Sutton	3.00	Defensive	3.7	18	19.22	0.26	0.26
Anton Volchenkov	2.50	Defensive	3.7	64	20.68	0.99	0.99	Anton Volchenkov	2.50	Defensive	3.7	64	20.68	0.99	0.99
Chris Campoli	0.63	Average	0.0	67	17.83	0.00	0.00	Chris Campoli	0.63	Average	0.0	67	17.83	0.00	0.00
Brian Lee	1.25	Average	0.0	23	15.55	0.00	0.00	Chris Phillips	3.50	Average	0.0	82	22.33	0.00	0.00
Chris Phillips	3.50	Average	0.0	82	22.33	0.00	0.00	Alexandre Picard	0.80	Average	0.0	45	19.04	0.00	0.00
Alexandre Picard	0.80	Average	0.0	45	19.04	0.00	0.00	Matt Carkner	0.50	Physical	0.0	53	16.49	0.00	0.00
Matt Carkner	0.50	Physical	0.0	51	16.90	0.00	0.00	Ladislav Smid	1.30	Physical	0.0	51	16.57	0.00	0.00
Defense Total	17.18	9	493	9,696.50	5.77	5.77		Defense Total	17.23	9	493	9,693.98	5.77	5.77	
Defense Max			492	9,840.00				Defense Max			492	9,840.00			
Goalies								Goalies							
Brian Elliott	0.85	Elite	32.1	48	63.29	19.63	19.63	Brian Elliott	0.85	Elite	32.1	48	63.29	19.63	19.63
Pascal Leclaire	3.80	Bottom	0.0	31	56.28	0.00	0.00	Pascal Leclaire	3.80	Bottom	0.0	31	56.28	0.00	0.00
Goalies Total	4.65	2	79	4,782.73	19.63	19.63		Goalies Total	4.65	2	79	4,782.73	19.63	19.63	
Goalies Max			82	4,920.00				Goalies Max			82	4,920.00			
Overall								Overall							
Team Total	61.77	26	1,521	28,741.38	82.81	82.81		Team Total	60.87	26	1,521	28,739.81	84.08	84.08	
Team Max			1,558	29,520.00				Team Max			1,558	29,520.00			
Trade Impact								Trade Impact							
								-0.90							
								1.27							
								1.27							

Figure C.2: The screenshot shows Ottawa's projected team compositions after Dany Heatley was traded to San Jose and Ottawa's projected team composition if Heatley had been traded to Edmonton.

Acknowledgments

The authors thank Birsen Donmez for helpful discussions on clustering and regression. The authors also thank an associate editor and two anonymous reviewers for their comments and suggestions, which significantly improved the quality of this paper.

References

- Anderberg, M. R. 1973. *Cluster Analysis for Applications*. Academic Press, New York.
- Beaudoin, D., T. B. Swartz. 2010. Strategies for pulling the goalie in hockey. *Amer. Statistician* 64(3) 197–204.
- Burger, J. D., S. J. K. Walters. 2003. Market size, pay, and performance: A general model and application to Major League Baseball. *J. Sports Econom.* 4(2) 108–125.
- Calinski, T., J. Harabasz. 1974. A dendrite method for cluster analysis. *Comm. Statist.* 3(1) 1–27.
- Cochran, J. J., R. Blackstock. 2009. Pythagoras and the National Hockey League. *J. Quant. Anal. Sports* 5(2) Article 11.
- Costa, D. 1995. An evolutionary tabu search algorithm and the NHL scheduling problem. *INFOR* 33(3) 161–178.
- Erkut, E. 1987. More on Morrison and Wheat's "Pulling the goalie revisited." *Interfaces* 17(5) 121–123.
- Ferland, J. A., C. Fleurent. 1991. Computer aided scheduling for a sport league. *INFOR* 29(1) 14–25.
- Fleisher, A. A., B. L. Goff, R. D. Tollison. 1992. *The National Collegiate Athletic Association: A Study in Cartel Behavior*. University of Chicago Press, Chicago.
- Fleurent, C., J. A. Ferland. 1993. Allocating games for the NHL using integer programming. *Oper. Res.* 41(4) 649–654.
- Gennaro, V. 2007. *Diamond Dollars: The Economics of Winning in Baseball*. Maple Street Press, Hanover, MA.
- Ingolfsson, A. 2004. Simulating NHL games to motivate student interest in OR/MS. *INFORMS Trans. Ed.* 5(1) 37–46.
- Ingolfsson, A. 2010. Ice hockey. *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Hoboken, NJ.
- Jones, J. C. H., S. Nadeau, W. D. Walsh. 1997. The wages of sin: Employment and salary effects of violence in the National Hockey League. *Atlantic Econom. J.* 25(2) 191–206.
- Jones, J. C. H., S. Nadeau, W. D. Walsh. 1999. Ethnicity, productivity and salary: Player compensation and discrimination in the National Hockey League. *Appl. Econom.* 31(5) 593–608.
- Lucifora, C., R. Simmons. 2003. Superstar effects in sport: Evidence from Italian soccer. *J. Sports Econom.* 4(1) 35–55.
- Morrison, D. G. 1976. On the optimal time to pull the goalie: A Poisson model applied to a common strategy used in ice hockey. *Management Science in Sports*, Vol. 4. TIMS Studies in Management Science. North Holland, Amsterdam.
- Morrison, D. G., R. D. Wheat. 1986. Misapplications reviews: Pulling the goalie revisited. *Interfaces* 16(6) 28–34.
- Mullet, G. M. 1977. Simeon Poisson and the National Hockey League. *Amer. Statistician* 31(1) 8–12.
- Nydick, R. L., Jr., H. J. Weiss. 1989. More on Erkut's "More on Morrison and Wheat's 'Pulling the goalie revisited.'" *Interfaces* 19(5) 45–48.
- Rousseeuw, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20(1) 73–85.
- Russell, T., P. vanBeek. 2009. Determining the number of games needed to guarantee an NHL playoff spot. *6th Internat. Conf. Integration AI OR Techniques Constraint Programming Combin. Optim. Problems, Pittsburgh*.
- Schmidt, M. B., D. J. Berri. 2001. Competitive balance and attendance: The case of Major League Baseball. *J. Sports Econom.* 2(2) 145–167.
- Thomas, A. C. 2006. The impact of puck possession and location on ice hockey strategy. *J. Quant. Anal. Sports* 2(1) Article 6.
- Vincent, C. B., B. Eastman. 2009. Defining the style of play in the NHL: An application of cluster analysis. *J. Quant. Anal. Sports* 5(1) Article 10.
- Washburn, A. 1991. Still more on pulling the goalie. *Interfaces* 21(2) 59–64.