NHL Shot Quality 2009-10

A look at shot angles and rebounds

Ken Krzywicki – June 2010

Overview

As in prior regular seasons, RTSS data was collected from the NHL in order to assess the quality of shots taken. However, after reading Gabe Desjardins' work using position on the ice, (x,y) coordinates, this additional data source was obtained from ESPN's website and appended to the NHL data in order to see what additional value, if any, it could bring to examining shot quality. A big "thanks" goes out to Gabe for helping me translate the ESPN data into English.

Data

As usual, RTSS data was collected from the NHL. As well, ESPN data, which contain the (x,y) coordinates for all events, was appended to it. Game 1163 ESPN data, however, was not available. The (x,y) coordinates represent the distance from the center ice red line towards the attacking goal; the distance from center ice to the goal line is 89 feet.

All regulation and overtime shots on goal were collected; empty net goals were excluded because they always go in. Since the data had issues, it was cleansed using the following three steps:

Step One: Shot Distance and Zone Reclassification

	Theoretical Distance	
	Min	Max
Offensive	0	77
Neutral	64	122
Defensive	115	189

There were instances where the zone was incorrectly labeled or the distance was not correct. The first step in data cleansing was to calculate the minimum and maximum distance that a shot from each zone could take (see table at right) and compare the distance reported with these theoretical values. If the shot was labeled as originating in the offensive zone and the distance was greater than 77 feet, the distance was recomputed using the Pythagorean formula as follows:

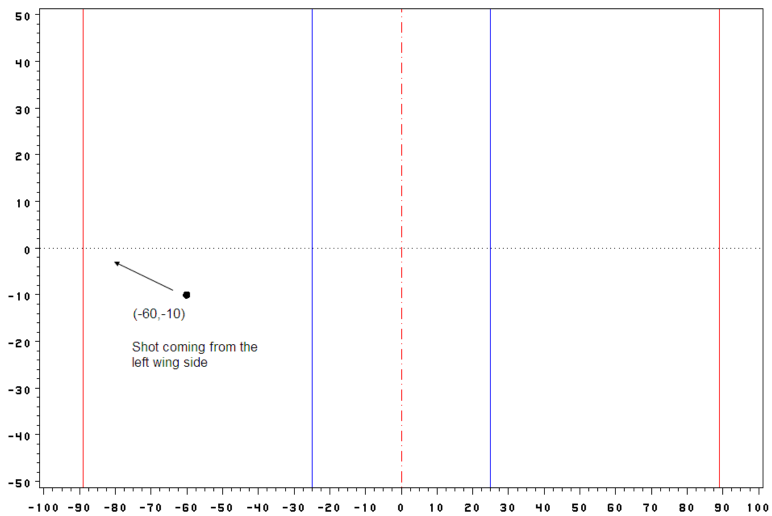
$$D_{Calc} = \sqrt{(89 - abs(x))^2 + y^2} \qquad (1)$$

If necessary, the zone was reclassified based on the theoretical minimum and maximum values given above.

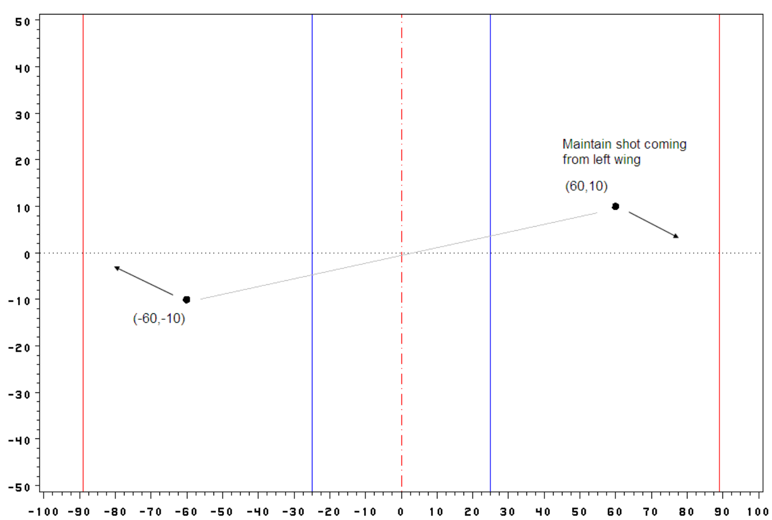
A similar cleansing for the neutral zone was also performed. There were 579 shots reportedly coming from the defensive zone, 32 of which were goals. Since empty net goals were ignored for this study, it was unlikely that goalies were beat 32 times from the shooter's defensive zone. Therefore, all shots listed as coming from the defensive zone were reclassified accordingly. I do realize that some shots on goal probably did originate in the defensive zone, but will accept the marginal error.

Step 2: X-Y Reclassification

Since teams shoot at different ends of the rink, the (x,y) coordinates will vary based on the team in question and the period. That is why equation (1) above used the absolute value of the x-coordinate when we derived the computed distance. The example shot shown at right is one coming off the left wing, with respect to the shooter. If we were to plot all the (x,y) shots for an entire game, we would see dots all over the graph. Therefore, I decided to transpose the (x,y) coordinates so they all faced the same ("east end") goal.



For shots in the offensive zone, this was trivial: I simply ensured that the x-coordinate was positive. If it was not, the sign of both the x- and y-coordinate was flipped. Using our example of a reported location of (-60,-10), a shot coming from the left wing, I reclassified that to read (60,10) to maintain not only proper zone, but from which wing the shot came.



Shots coming from the neutral zone required a little more work and thought. Those from the attacking side of the red line should have a positive x-coordinate; the opposite for those from the defensive side of the red line. In order to come up with a set of rules for flipping or not flipping the (x,y) coordinates, I assumed that a majority of a team's neutral zone shots came from the offensive side of the red line. For each team, for each period of each game, I computed the percentage of shots with positive and negative x-coordinates. If a majority were positive, no (x,y) sign flipping was required; otherwise the signs were changed.

For both offensive and neutral zone shots, about half each had their (x,y) signs flipped. This makes intuitive sense based on the fact that each team faces each side of the rink roughly an equal number of times.

There were also a dozen instances where the NHL and ESPN distances did not match. These were mostly due to the NHL (correctly) changing the scoring after the fact; the ESPN data showed a long shot and the NHL data showed a deflection or tip-in from a closer distance with the ESPN shooter receiving the primary assist. Since the ESPN data was not corrected with respect to who scored the goal, I assumed the (x,y) coordinates also incorrect and set them to missing.

Step 3: Remove Distance Bias

The under- and over-reporting of *distance* across the NHL rinks is a well-known phenomenon[1]. To remove this bias, a general linear model (least squares) controlling for rink to predict the reported distance was built and the residuals of that model, which I'm calling *adjusted distance*, were used in a shot quality model. The correlation between reported and adjusted distance was about 86%, so I opted to use the adjusted value as in the past. Recall that an *adjusted distance* of, say, -25.32 represents a shot reported roughly 25 feet less than the reported rink mean distance.

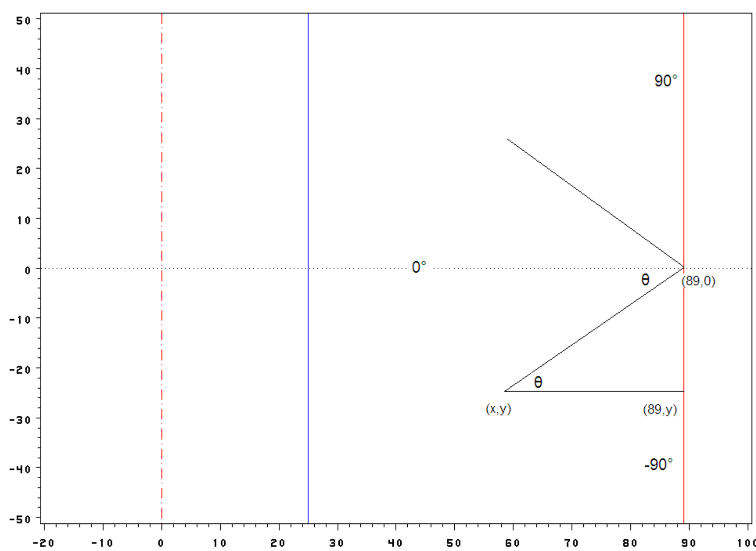
One might argue that the (x,y) coordinates are also biased by rink. However, if distance is incorrect that does not necessarily mean that (x,y) are also incorrect. A shot reported from 30 feet at a 20° angle might really be a 35-foot shot from the same angle. And it is the angle of the shot

[1] Please refer to *Product Recall Notice for Shot Quality*, June 2007 [Ryder] and *Removing Observer Bias from Shot Distance*, September 2009 [Krzywicki]

that was considered for inclusion in the shot quality model. For comparison sake, I did calculate an adjusted angle[2] and the correlation between it and the reported angle was 99%; therefore the angle is used as-is in model development since it is easier to interpret and visualize.

Examining the Angles

Where the (x,y) was not missing or where $x \neq 89$, the angle of the shot taken relative to the goalie's perspective was calculated (see right and equation (2) below). Missing angles accounted for only 0.2% of the shots on goal, an immaterial matter.



$$\theta = Tan^{-1}\left(\frac{y}{89-x}\right)\left(\frac{180}{\pi}\right) \qquad (2)$$

The goal rates were rather symmetrical with respect to shot angle, with the most dangerous shots coming from straight on, where there is the most net to see (2.4% of all shots). As well, those taken within 1° to 15° (either side) of the goaltenders were considerably more dangerous and comprised 22.0% of all shots. Defensemen try to take the opponent wide … and for good reason. This intuitive stance that there is less net to see from "bad" angles is backed up by the numbers. It is

Shot Angle (deg)		Shots	Pct of Shots	Goals	Pct of Goals	Goal Rate
Missing		142	0.2%	31	0.5%	21.8%
-90 - -46	Right Wing	8,490	11.4%	535	8.1%	6.3%
-45 - -31	Right Wing	8,962	12.0%	682	10.3%	7.6%
-30 - -16	Right Wing	11,272	15.2%	962	14.6%	8.5%
-15 - -1	Right Wing	8,086	10.9%	986	14.9%	12.2%
0		1,769	2.4%	330	5.0%	18.7%
1 - 15	Left Wing	8,260	11.1%	1,023	15.5%	12.4%
16 - 30	Left Wing	11,580	15.6%	948	14.4%	8.2%
31 - 45	Left Wing	8,974	12.1%	671	10.2%	7.5%
46 - 90	Left Wing	6,842	9.2%	431	6.5%	6.3%
Total		74,377	100.0%	6,599	100.0%	8.9%

[2] GLM method, as with *distance*, was employed.

interesting, though, to note that angles from 15° to 30° (on either side of the net) are near the average goal rate of 8.9 percent. Taking an opponent wider than 30° on either side gives the goalie much easier shots to handle—yet only 35% of all goals came from angles greater than 30 degrees. This bit of information may factor into a team's defensive system.

For modeling, the absolute value of the shot angle was used in a continuous fashion, running from zero to ninety degrees; missing angles were assigned a value of zero since we cannot quantify the extent of their quality.

Rebounds in Depth

We have seen numerous times in the past that rebound shots are more dangerous than non-rebound shots. The rebound variable has been in all of my shot quality models as well as those of other analysts. The goal rate for rebound shots of 32.8% is 3.7 times higher than the overall goal rate of 8.9%, indicating their high quality. Rebounds constituted 4% of all shots taken.

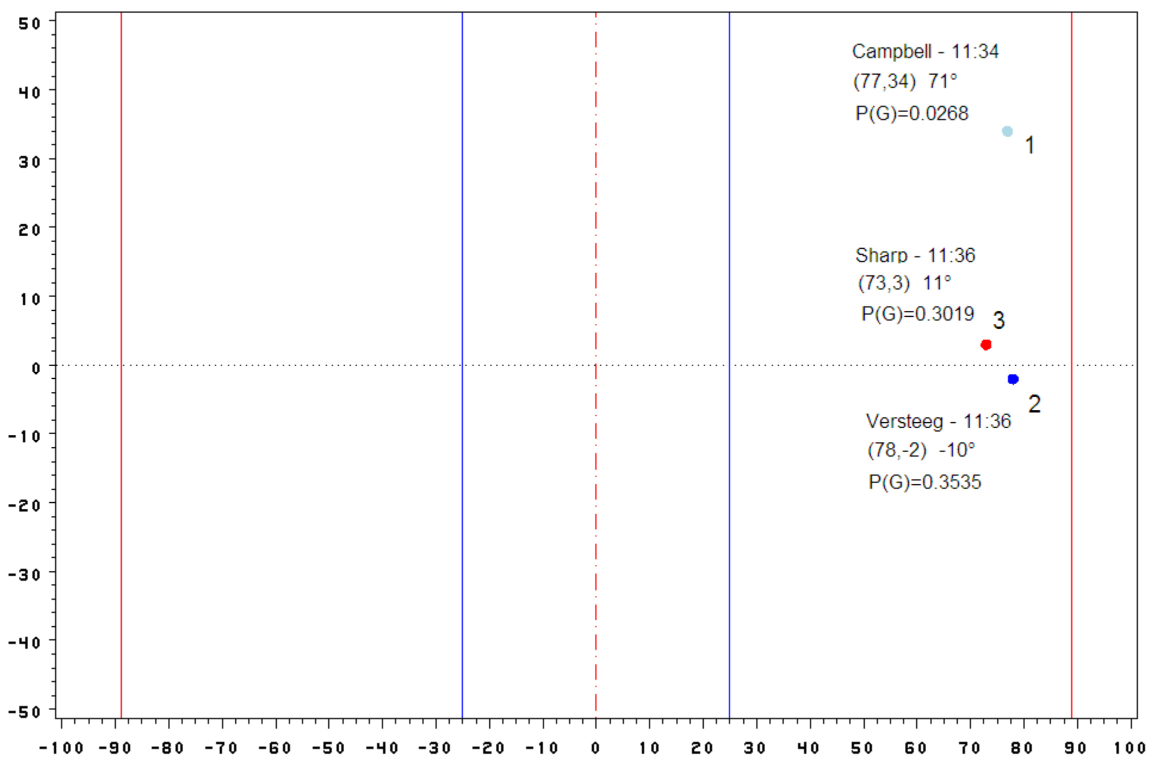
Push Direction	Rebound Shots	Pct	Goals	Goal Rate
Missing	17	0.6%	5	29.4%
No push	61	2.0%	10	16.4%
Left	1,505	50.1%	482	32.0%
Right	1,422	47.3%	488	34.3%
Total	3,005	100.0%	985	32.8%

With the addition of shot angle data, another variable was created for rebounds—the "push" a goalie had to make to face the rebound shot, given the angle of the prior (rebound generating) shot, i.e., the difference between the two angles. Values greater than zero represent a push to the right (blocker side for a majority, though not all, goalies), while those less than zero represent a push to the left (glove side for most). A rebound shot coming at exactly the same angle as the rebound generating shot required no push.

Push Degrees		Rebound Shots	Pct	Goals	Goal Rate
Missing		17	0.6%	5	29.4%
-180 to -31	Left	688	22.9%	267	38.8%
-30 to -16		325	10.8%	99	30.5%
-15 to 15		1,005	33.4%	247	24.6%
16 to 30	Right	351	11.7%	123	35.0%
31 to 180		619	20.6%	244	39.4%
Total		3,005		985	32.8%

The near-perfect symmetry we saw for the shot angle did not quite hold up for degrees a goalie had to push. We note from the table that pushing off to the right to save a rebound shot was slightly more difficult. Perhaps this is due to the fact that most goalies catch with their left hand and a blocker save offers a more difficult situation. We see a slightly higher goal rate pushing to the right 31 or more degrees than we do the same amount to the left. The interesting comparison is between moving 16° to 30° to the right versus the same angle to the left: 35.0% versus 30.5%, each with roughly 11% of all rebound shots.

One example of a goalie having to move quite a bit in a short period of time comes from game seven—Chicago at Florida—with Tomas Vokoun, one of the few goalies to catch with his right hand, in net. The first shot from Brian Campbell came at 11:34 of the third period. Vokoun handled his shot from an angle of 71° and even pushed 81° to his left (blocker side) to save Kris Versteeg's wrist shot from an angle of -10 degrees. We could say that was a great save due to the high probability, 0.3535, of going in. However, Vokoun was not able to push back another 21° to the right (glove side) in less than a second to stop Patrick Sharp's wrist shot, which may be thought of as a "double rebound."



Video found at the NHL website revealed that Vokoun was down and out after the Versteeg shot; had the defense been able to clear that rebound, Sharp may not have had such an easy goal. Video also revealed that the RTSS mislabeled Campbell's shot as a backhand; it was in fact a wrist shot. However, this has no bearing on the shot quality model since both types of shots receive that same weight (see model below).

Though the difference in angles from the rebound generating shot to the rebound shot is an interesting variable, it did not make it to the final model due to high correlation to the more powerful *rebound* variable. Still, I presented it above due to intrinsic interest.

Model Methodology

Using a 70% random sample of the 2009-10 NHL play-by-play (PBP) files from the RTSS scoring system and (x,y) coordinates obtained from the ESPN website, a logistic regression was run. Note that the shot angle was not forced into the model, rather entered of its own accord and proved rather significant. Conspicuously missing are most shot types. Other than wrap-around, *shot type* did not make it into the model. I surmise that the shot angle provided a better predictor in the multivariate sense of the model, thus keeping *shot type* mostly out. As well, the shot angle helped temper the distance variable: in the past, a shorter shot from any angle was given the same weight; now those short shots from "bad" angles are differentiated from short shots at "good" angles.

Model Results

The logistic regression model given below was built on a 70% random sample of the shots on goal, excluding empty net goals, and validated on the 30% held out. The model did not over fit the training dataset and was deemed appropriate.

Variable	Interval	Points	Marginal Contribution
Intercept	Add to all records	-2.2899	
Distance (adjusted)	Multiply points by adjusted distance	-0.0437	0.074
Shot angle	Multiply points by abs(shot angle)	-0.0162	0.016
Rebound	No	0.0000	0.013
	Yes	0.9948	
Situation (shooting team)	EV	0.0000	0.005
	SH	0.0000	
	PP	0.4370	
Shot after give-away (opposing team)	No	0.0000	0.002
	Yes	0.5623	
Shot Type	Wrap	-1.0379	0.002
	Other	0.0000	

Shot quality is defined as:

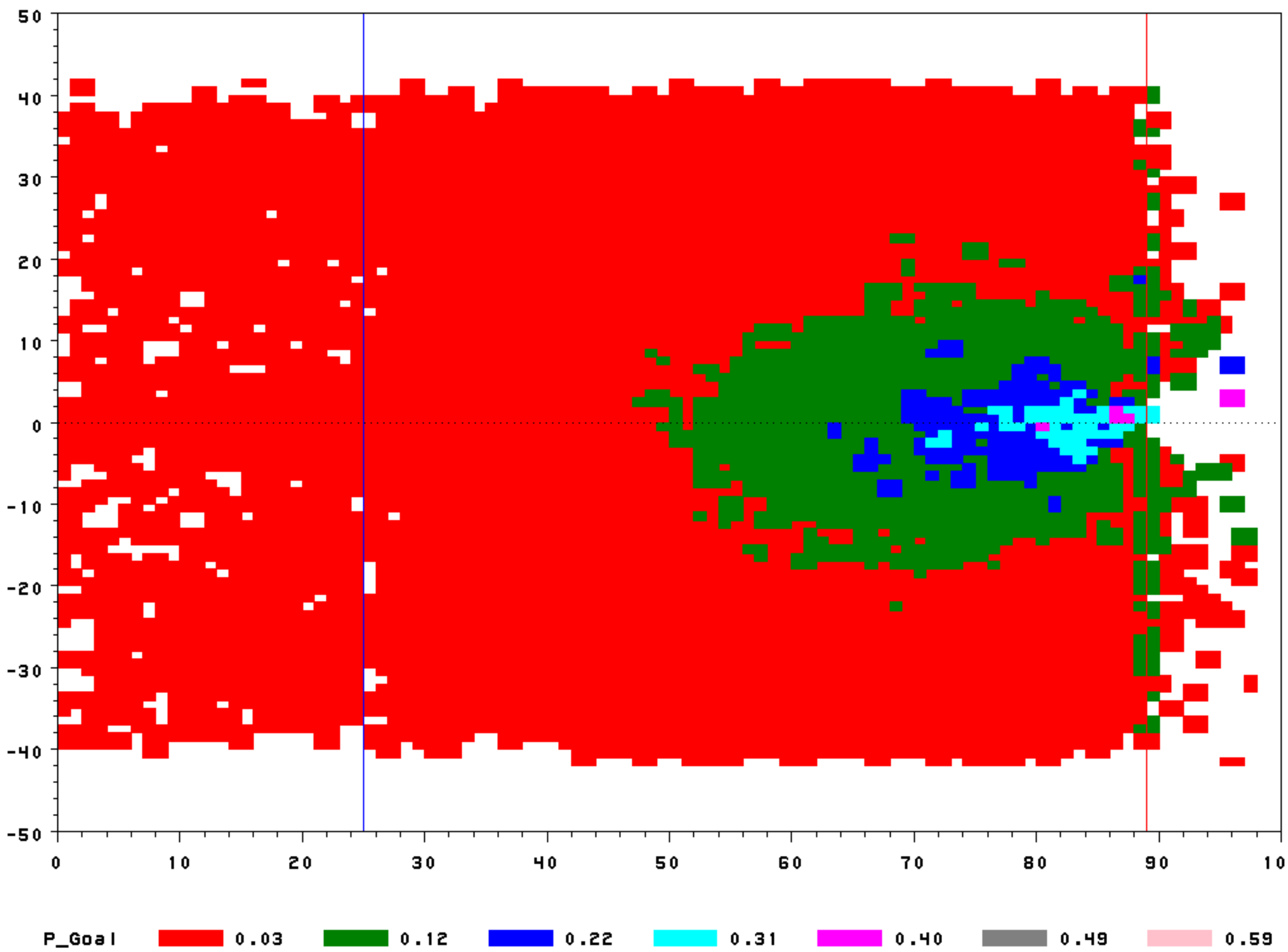
$$P(\mathrm{GOAL}) = \frac{1}{1+e^{-\sum \mathrm{points}}}$$

Applying the model to the entire population we find that the model fit the data very well:

KS = 36.67 IV = 79.31 D = 0.812 ROC = 0.740 Gini = 0.480

Score	Range	Totals	Cuml %	Save	Int Rate	Cuml %	Goal	Int Rate	Cuml %	Avg Scr
0.185654	0.616943	7,437	10.00%	5,479	73.67%	8.08%	1,958	26.33%	29.67%	0.270407
0.136197	0.185638	7,438	20.00%	6,193	83.26%	17.22%	1,245	16.74%	48.54%	0.158609
0.103556	0.136183	7,438	30.00%	6,474	87.04%	26.77%	964	12.96%	63.15%	0.118869
0.078872	0.103554	7,438	40.00%	6,760	90.88%	36.75%	678	9.12%	73.42%	0.090580
0.061006	0.078862	7,437	50.00%	6,913	92.95%	46.95%	524	7.05%	81.36%	0.069429
0.048719	0.061005	7,438	60.00%	7,043	94.69%	57.34%	395	5.31%	87.35%	0.054662
0.038769	0.048718	7,438	70.00%	7,132	95.89%	67.86%	306	4.11%	91.98%	0.043534
0.030631	0.038764	7,438	80.00%	7,220	97.07%	78.51%	218	2.93%	95.29%	0.034566
0.023716	0.030630	7,438	90.00%	7,266	97.69%	89.23%	172	2.31%	97.89%	0.027166
0.001670	0.023715	7,437	100.00%	7,298	98.13%	100.00%	139	1.87%	100.00%	0.017887
Total		74,377		67,778	91.13%		6,599	8.87%		0.088570

A contour plot of the model's predicted P(GOAL) relative to the (x,y) data and viewed from the attacking side of the center red line ($x = 0$) looks as follows :



Conclusions

With an additional data source we were able to create additional variables for consideration when building a shot quality model, namely: shot angle and rebound push direction and degrees. Even though only one of these three made it into the model—possibly at the expense of the more traditional *shot type* variables—they all provided interesting insight into shot quality. Examining rebounds taught us that it was a bit more difficult to push off 16 or more degrees to the right than the same magnitude to the left.

As in the past, we were able to construct a robust model to describe the shots taken in the prior season and apply them at the individual level in order to compare actual versus predicted performance—a valuable player and team evaluation tool.

Appendix – Team by Team Comparison

Below find a table of SQA = 1 – P(GOAL) and SQF = P(GOAL) sorted by SQA.

	Team	SA	GA	SvPct	SQA		SF	GF	ShotPct	SQF
	TOTAL	74,377	6,599	.911	.911		74,377	6,599	8.9	8.9
1	CBJ	2,512	246	.902	.917	Easier SQA	2,327	203	8.7	8.6
2	MIN	2,406	230	.904	.917		2,263	211	9.3	9.1
3	NYR	2,440	204	.916	.916		2,418	212	8.8	8.2
4	NJD	2,211	184	.917	.915		2,446	203	8.3	8.7
5	COL	2,619	218	.917	.914		2,280	229	10.0	9.5
6	CGY	2,367	198	.916	.914		2,343	193	8.2	8.3
7	DET	2,410	204	.915	.913		2,727	213	7.8	8.7
8	NAS	2,389	214	.910	.913		2,501	210	8.4	8.5
9	DAL	2,551	235	.908	.913		2,520	226	9.0	9.2
10	FLA	2,790	222	.920	.913		2,323	194	8.4	8.8
11	STL	2,464	208	.916	.913		2,440	213	8.7	8.4
12	PHO	2,420	189	.922	.912		2,498	205	8.2	8.6
13	LAK	2,256	206	.909	.912		2,381	223	9.4	8.7
14	MTL	2,630	214	.919	.912		2,339	203	8.7	9.1
15	ANA	2,728	234	.914	.911	Avg SQA	2,465	226	9.2	8.8
16	PHI	2,343	217	.907	.911		2,582	225	8.7	9.1
17	SJS	2,567	205	.920	.911		2,596	249	9.6	9.4
18	NYI	2,615	254	.903	.911		2,490	210	8.4	9.2
19	WAS	2,534	226	.911	.910	Tougher SQA	2,682	302	11.3	9.3
20	EDM	2,710	271	.900	.910		2,319	204	8.8	9.1
21	BOS	2,442	188	.923	.910		2,595	192	7.4	8.6
22	BUF	2,567	194	.924	.910		2,585	228	8.8	9.4
23	PIT	2,340	225	.904	.910		2,683	243	9.1	8.8
24	OTT	2,333	226	.903	.910		2,429	211	8.7	9.0
25	VAN	2,412	211	.913	.910		2,523	257	10.2	8.8
26	TOR	2,430	252	.896	.910		2,668	207	7.8	8.7
27	CAR	2,559	236	.908	.909		2,390	214	9.0	9.0
28	CHI	2,054	199	.903	.908		2,797	260	9.3	8.4
29	ATL	2,709	242	.911	.904		2,405	225	9.4	8.9
30	TBL	2,569	247	.904	.904		2,362	208	8.8	9.1

Appendix – Goalie Comparison

GOALIE	TEAM	SA	GA	SvPct	SQA
TOTAL	NHL	74,377	6,599	.911	.911
GIGUERE	ANA	580	58	.900	.911
HILLER	ANA	1,860	152	.918	.911
MCELHINNEY	ANA	288	24	.917	.913
Total	ANA	2,728	234	.914	.911
HEDBERG	ATL	1,356	115	.915	.905
PAVELEC	ATL	1,353	127	.906	.904
Total	ATL	2,709	242	.911	.904
RASK	BOS	1,221	84	.931	.907
THOMAS	BOS	1,221	104	.915	.914
Total	BOS	2,442	188	.923	.910
ENROTH	BUF	37	4	.892	.908
LALIME	BUF	432	40	.907	.908
MILLER	BUF	2,098	150	.929	.910
Total	BUF	2,567	194	.924	.910
LEGACE	CAR	745	69	.907	.912
LEIGHTON	CAR	164	25	.848	.903
PETERS	CAR	241	23	.905	.906
WARD	CAR	1,409	119	.916	.908
Total	CAR	2,559	236	.908	.909
GARON	CBJ	858	83	.903	.916
MASON	CBJ	1,654	163	.901	.918
Total	CBJ	2,512	246	.902	.917
KIPRUSOFF	CGY	2,035	163	.920	.913
MCELHINNEY	CGY	235	27	.885	.916
TOSKALA	CGY	97	8	.918	.923
Total	CGY	2,367	198	.916	.914
CRAWFORD	CHI	35	3	.914	.921
HUET	CHI	1,083	114	.895	.910
NIEMI	CHI	936	82	.912	.906
Total	CHI	2,054	199	.903	.908
ANDERSON	COL	2,233	186	.917	.914
BUDAJ	COL	386	32	.917	.917
Total	COL	2,619	218	.917	.914

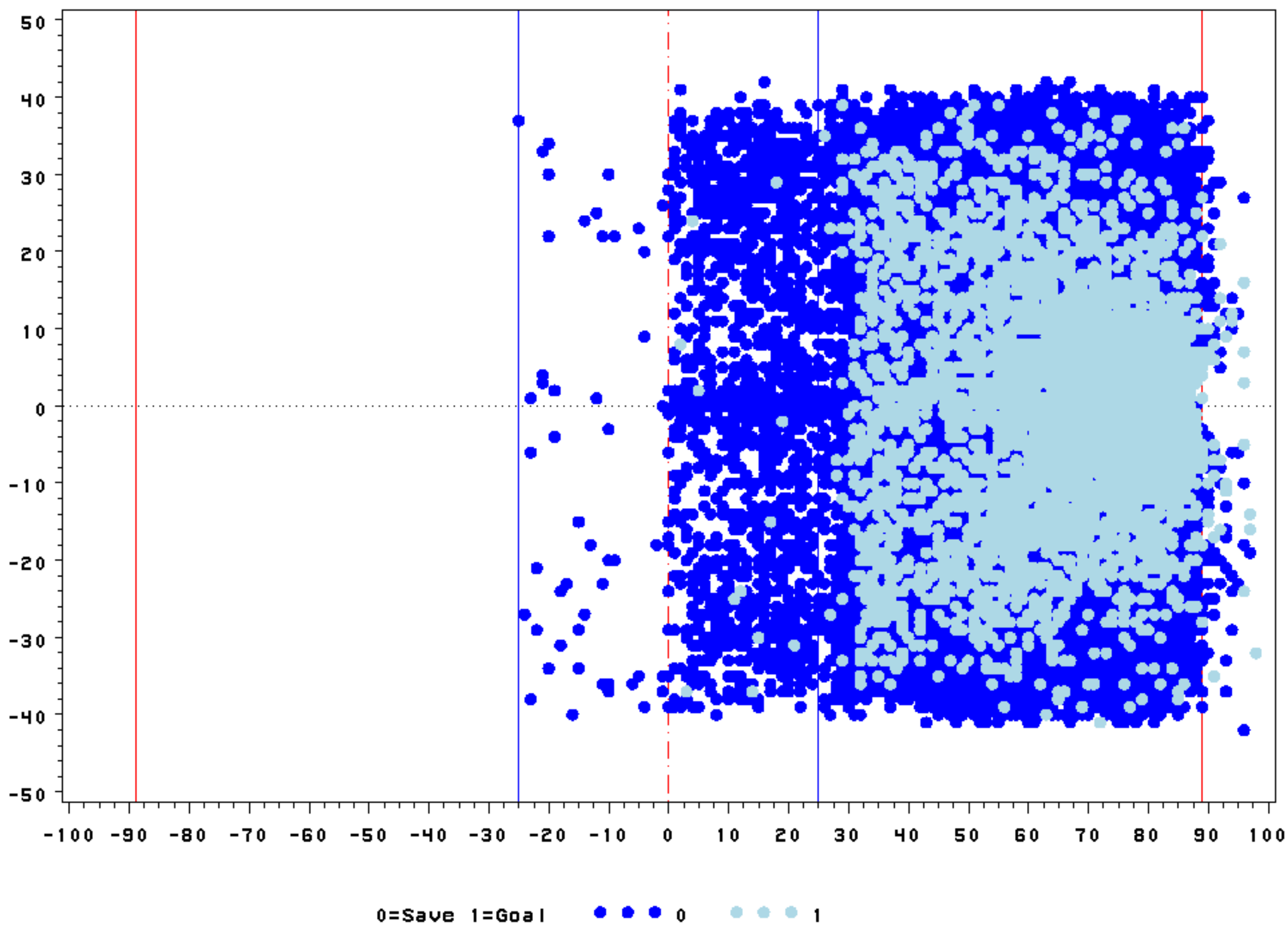
GOALIE	TEAM	SA	GA	SvPct	SQA
TOTAL	NHL	74,377	6,599	.911	.911
AULD	DAL	558	59	.894	.918
CLIMIE	DAL	38	5	.868	.908
LEHTONEN	DAL	350	31	.911	.906
TURCO	DAL	1,605	140	.913	.913
Total	DAL	2,551	235	.908	.913
HOWARD	DET	1,849	141	.924	.915
OSGOOD	DET	561	63	.888	.908
Total	DET	2,410	204	.915	.913
DESLAURIERS	EDM	1,529	152	.901	.908
DUBNYK	EDM	579	64	.889	.911
KHABIBULIN	EDM	602	55	.909	.914
Total	EDM	2,710	271	.900	.910
CLEMMENSEN	FLA	668	59	.912	.913
SALAK	FLA	40	6	.850	.902
VOKOUN	FLA	2,082	157	.925	.913
Total	FLA	2,790	222	.920	.913
BERNIER	LAK	94	4	.957	.922
ERSBERG	LAK	234	22	.906	.911
QUICK	LAK	1,928	180	.907	.911
Total	LAK	2,256	206	.909	.912
BACKSTROM	MIN	1,632	158	.903	.918
DUBIELEWICZ	MIN	34	5	.853	.928
HARDING	MIN	692	66	.905	.912
KHUDOBIN	MIN	48	1	.979	.915
Total	MIN	2,406	230	.904	.917
HALAK	MTL	1,386	105	.924	.910
PRICE	MTL	1,244	109	.912	.914
Total	MTL	2,630	214	.919	.912
ELLIS	NAS	848	77	.909	.912
RINNE	NAS	1,541	137	.911	.914
Total	NAS	2,389	214	.910	.913
BRODEUR	NJD	2,004	168	.916	.915
DANIS	NJD	207	16	.923	.914
Total	NJD	2,211	184	.917	.915

GOALIE	TEAM	SA	GA	SvPct	SQA
TOTAL	NHL	74,377	6,599	.911	.911
BIRON	NYI	859	89	.896	.910
DIPIETRO	NYI	201	20	.900	.926
ROLOSON	NYI	1,555	145	.907	.910
Total	NYI	2,615	254	.903	.911
AULD	NYR	52	5	.904	.920
JOHNSON	NYR	135	11	.919	.916
LUNDQVIST	NYR	2,109	167	.921	.917
VALIQUETTE	NYR	128	19	.852	.903
ZABA	NYR	16	2	.875	.910
Total	NYR	2,440	204	.916	.916
BRODEUR	OTT	87	3	.966	.914
ELLIOTT	OTT	1,424	130	.909	.910
LECLAIRE	OTT	822	93	.887	.908
Total	OTT	2,333	226	.903	.910
BACKLUND	PHI	24	2	.917	.905
BOUCHER	PHI	796	80	.899	.915
DUCHESNE	PHI	4	1	.750	.929
EMERY	PHI	784	74	.906	.907
LEIGHTON	PHI	735	60	.918	.912
Total	PHI	2,343	217	.907	.911
BRYZGALOV	PHO	1,961	156	.920	.912
LABARBERA	PHO	459	33	.928	.913
Total	PHO	2,420	189	.922	.912
CURRY	PIT	14	5	.643	.882
FLEURY	PIT	1,772	168	.905	.910
JOHNSON	PIT	541	51	.906	.910
PECHURSKI	PIT	13	1	.923	.915
Total	PIT	2,340	225	.904	.910
GREISS	SJS	399	35	.912	.905
NABOKOV	SJS	2,168	170	.922	.912
Total	SJS	2,567	205	.920	.911
CONKLIN	STL	764	60	.921	.915
MASON	STL	1,700	148	.913	.912
Total	STL	2,464	208	.916	.913

GOALIE	TEAM	SA	GA	SvPct	SQA
TOTAL	NHL	74,377	6,599	.911	.911
NIITTYMAKI	TBL	1,388	127	.909	.902
SMITH	TBL	1,165	117	.900	.906
TOKARSKI	TBL	16	3	.813	.922
Total	TBL	2,569	247	.904	.904
GIGUERE	TOR	451	38	.916	.910
GUSTAVSSON	TOR	1,146	112	.902	.911
MACDONALD	TOR	157	17	.892	.917
TOSKALA	TOR	676	85	.874	.906
Total	TOR	2,430	252	.896	.910
LUONGO	VAN	1,915	167	.913	.910
RAYCROFT	VAN	438	39	.911	.911
SCHNEIDER	VAN	59	5	.915	.902
Total	VAN	2,412	211	.913	.910
NEUVIRTH	WAS	464	40	.914	.913
THEODORE	WAS	1,352	121	.911	.905
VARLAMOV	WAS	718	65	.909	.919
Total	WAS	2,534	226	.911	.910

Appendix – Shots on Goal Chart

With the addition of (x,y) coordinates we are able to visually depict each shot on goal. After the data cleansing discussed above, here is what all 1,230 games combined looked like (74,377 shots on goal). Recall that these are for periods 1 through 4 and exclude empty net goals.



Author Contact: kenkrzywicki@hotmail.com

Acknowledgements: I wish to thank Alan, Gabe, Carol, Corey, MJD and Denis.

For more information on hockey analytics, see Alan Ryder's web site: www.hockeyanalytics.com