Threat

Micah Blake McCurdy
hockeyviz.com

Vancouver, BC
Vancouver Hockey Analytics Conference
March 3, 2018

Introduction

Tired:	Numbers
Wired:	Pictures

Introduction

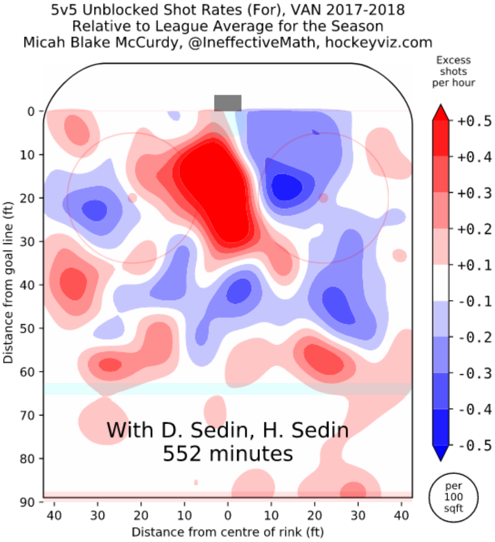
Tired:	Numbers
Wired:	Pictures

Vancouver with the Sedins at 5v5 generate 45 unblocked shots per hour, 5% more than league average.

Introduction

Tired: Numbers

Wired: Pictures



Aim

Isolate individual skater impact on shots, both for and against.

New Thing

Treat maps as first-class objects, instead of single-numbers like rates or counts.

Isolation

Control for the most important aspects of play which are *outside* of a player's control:

- Linemates
- Zone usage
- The score (!)
- Competition faced.

Isolation

Control for the most important aspects of play which are *outside* of a player's control:

- Linemates
- Zone usage
- The score (!)
- Competition faced. (Not yet, ask me later)

Least-Squares Regression

- α a collection of observations
- X a design matrix
- β a collection of (imagined) individual isolated impacts

$$X\beta = \alpha$$

Least-Squares Regression

- α a collection of observations
- X a design matrix
- β a collection of (imagined) individual isolated impacts

$$
\begin{aligned}
X\beta &= \alpha \\
X^T X\beta &= X^T\alpha
\end{aligned}
$$

Least-Squares Regression

- α a collection of observations
- X a design matrix
- β a collection of (imagined) individual isolated impacts

$$\begin{aligned} X\beta &= \alpha \\ X^T X\beta &= X^T\alpha \\ \beta &= (X^T X)^{-1} X^T\alpha \end{aligned}$$

Least-Squares Regression

- α a collection of observations
- X a design matrix
- β a collection of (imagined) individual isolated impacts

$$
\begin{aligned}
X\beta &= \alpha \\
X^T X\beta &= X^T\alpha \\
\beta &= (X^T X)^{-1} X^T \alpha
\end{aligned}
$$

Elements of α and β can be taken from *any inner product space* and the usual proof goes through.

Dreamy Wishful Thinking Interlude

What if we had observations α from *every* possible combination of k players from our team of n? What would that get us?

It would make X very simple:

$$X_{jp} = \begin{cases} \frac{1}{k} & \text{if } p \text{ is in the } j\text{-th } k\text{-combination of } n \\ 0 & \text{otherwise} \end{cases}$$

Closed form solutions

Commonly, least squares solutions of $X\beta = \alpha$ are obtained by:

- Minimizing $||X\beta - \alpha||$ with a fancy optimiser.
- Numerically computing $\beta = (X^T X)^{-1} X^T \alpha$ with clever linear algebra.

However, in our case, (because X is highly structured) we can work it out by hand:

$$\beta = (X^T X)^{-1} X^T \alpha$$

Montage with bad music (1/3)

$$\left(X^T X\right)_{pq} = \begin{cases} \frac{1}{k^2}\binom{n-1}{k-1} & \text{if } p = q \\ \frac{1}{k^2}\binom{n-2}{k-2} & \text{if } p \neq q \end{cases} = \begin{cases} \frac{1}{kn}\binom{n}{k} & \text{if } p = q \\ \frac{k-1}{kn(n-1)}\binom{n}{k} & \text{if } p \neq q \end{cases}$$

$$\left(X^T X\right)^{-1}_{pq} = \begin{cases} n\frac{(n-1)k+1-k}{n-k}\binom{n}{k}^{-1} & \text{if } p = q \\ n\frac{1-k}{n-k}\binom{n}{k}^{-1} & \text{if } p \neq q \end{cases}$$

Montage with bad music (2/3)

$$
\begin{aligned}
\beta_p &= \left[\left(X^TX\right)^{-1}X^T\alpha\right]_p = \sum_j \left[\left(X^TX\right)^{-1}X^T\right]_{pj}\alpha_j \\
&= \sum_j\sum_q \left(X^TX\right)^{-1}_{pq} X^T_{qj}\alpha_j = \sum_j\sum_q \left(X^TX\right)^{-1}_{pq} X_{jq}\alpha_j \\
&= \sum_j \sum_{q \text{ in } j} \left(X^TX\right)^{-1}_{pq}\frac{1}{k}\alpha_j \\
&= \sum_{j \text{ with } p}\ \sum_{q \text{ in } j} \left(X^TX\right)^{-1}_{pq}\frac{1}{k}\alpha_j + \sum_{j \text{ without } p}\ \sum_{q \text{ in } j} \left(X^TX\right)^{-1}_{pq}\frac{1}{k}\alpha_j \\
&= \frac{(k-1)e+d}{k}\sum_{j \text{ with } p}\alpha_j + e\sum_{j \text{ without } p}\alpha_j \\
&= \frac{(k-1)e+d}{k}\binom{n-1}{k-1}\overline{\alpha}_{j \text{ with } p} + e\binom{n-1}{k}\overline{\alpha}_{j \text{ without } p}
\end{aligned}
$$

Montage with bad music (3/3)

$$
\begin{aligned}
\frac{(k-1)e+d}{k}\binom{n-1}{k-1} &= \frac{1}{k}\frac{k}{n}\binom{n}{k}\left[(k-1)e+d\right] \\
&= \frac{1}{n}\binom{n}{k}\left[(k-1)n\frac{1-k}{n-k}\binom{n}{k}^{-1}\right. \\
&\quad \left. +n\frac{(n-1)k+1-k}{n-k}\binom{n}{k}^{-1}\right] \\
&= (k-1)\frac{1-k}{n-k}+\frac{(n-1)k+1-k}{n-k} \\
&= \frac{(k-1)(1-k)+(n-1)k+1-k}{n-k} \\
&= \frac{k(1-k)+(n-1)k}{n-k}=\frac{k(n-k)}{n-k} \\
&= k
\end{aligned}
$$

Montage with bad music (4/3)

$$
\begin{aligned}
e\binom{n-1}{k} &= \binom{n-1}{k} n \frac{1-k}{n-k} \binom{n}{k}^{-1} \\
&= \frac{(n-1)!nk!(n-k)!}{k!(n+k-1)!n!(n-k)}(1-k) \\
&= 1-k
\end{aligned}
$$

just as desired.

Closed form solutions

$$\beta_p = k \cdot \begin{pmatrix} \text{Average of} \\ \text{Entries in } \alpha \\ \text{With } p \end{pmatrix} - (k-1) \cdot \begin{pmatrix} \text{Average of} \\ \text{Entries in } \alpha \\ \text{Without } p \end{pmatrix}$$

Closed form solutions

$$\beta_p = 5\begin{pmatrix}\text{Average of}\\ \text{Entries in } \alpha\\ \text{With } p\end{pmatrix} - 4\begin{pmatrix}\text{Average of}\\ \text{Entries in } \alpha\\ \text{Without } p\end{pmatrix}$$

Closed form solutions

$$\beta_p = 3.1 \begin{pmatrix} \text{Average of} \\ \text{Entries in } \alpha \\ \text{With } p \end{pmatrix} - 2.2 \begin{pmatrix} \text{Average of} \\ \text{Entries in } \alpha \\ \text{Without } p \end{pmatrix}$$

Ridge regression with $\lambda = 0.5$. This requires standardization.

Observations

We need observations for *every* combination of 5 players.

Observations

We need observations for *every* combination of 5 players.

When observations aren't available, impute them:

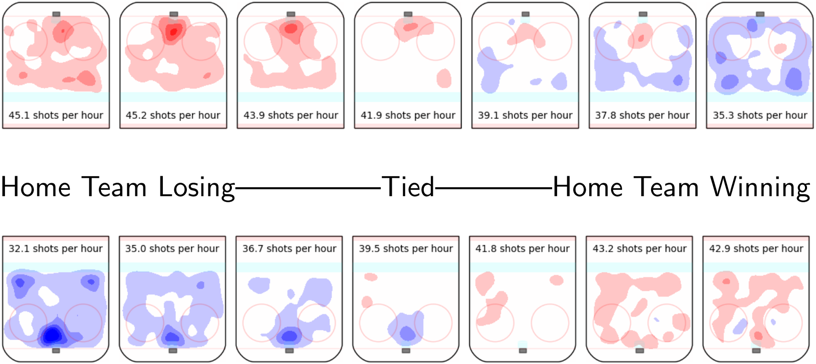
- by amalgamating all possible subsets of four players,
- by amalgamating all possible subsets of three,
- &c.

By hook or by crook, manufacture an α for every combination.

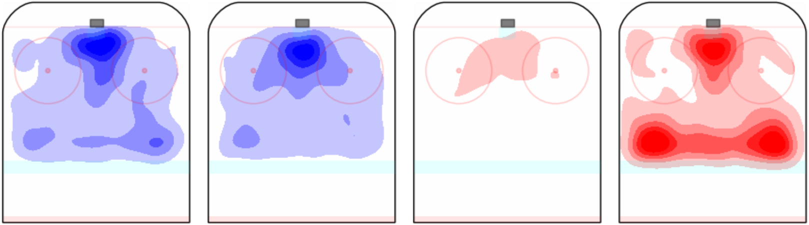
Adjustments

The regression adjusts for teammates, but not for score or zone usage. To account for them, we adjust each α for these things individually.

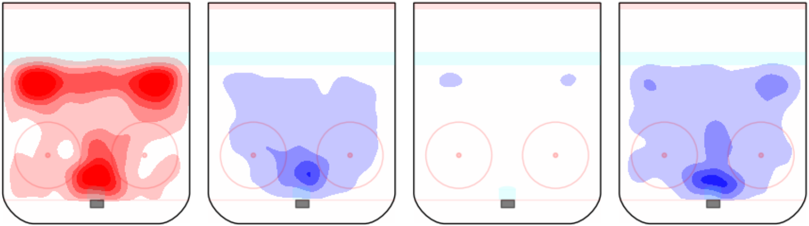
Score Adjustment



Zone Adjustment



Defensive Zone—Neutral Zone——On The Fly——Offensive Zone



Momentary Pausing to Consolidate Ground

- Treat shot densities as first-class objects.
- Adjust observations of a given set of players for score and zone.
- Impute observations for combinations who didn't play together.
- Use ridge regression to isolate individual performances.

Verification

- Consolidate impact into convenient units.
- Measure correlation from season to season. (for serious)
- Look at some interesting examples. (for insight)
- Look at some tails of the distribution this year. (for laughs)

Threat

To compare different players we weight their isolated shot contributions according to league average shooting percentages from given locations to obtain *threat*.

Threat

To compare different players we weight their isolated shot contributions according to league average shooting percentages from given locations to obtain *threat*.

Carefully avoiding shooting talent and goaltender talent.

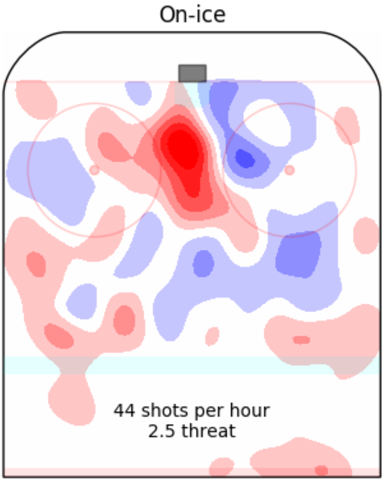
Threat

To compare different players we weight their isolated shot contributions according to league average shooting percentages from given locations to obtain *threat*.

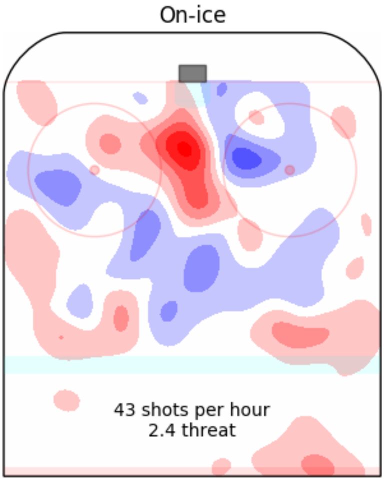
Carefully avoiding shooting talent and goaltender talent.

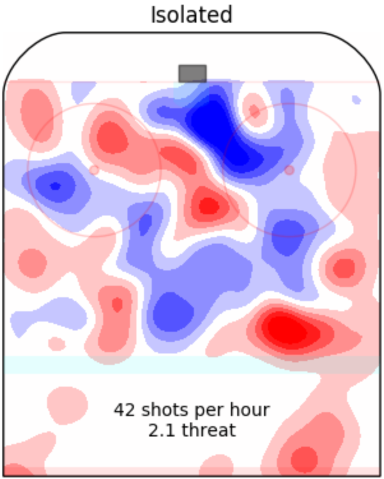
- Units of threat are goals per hour, purely from individual impact on shot locations.

2017-2018 Daniel Sedin, Observed vs Isolated, Offence

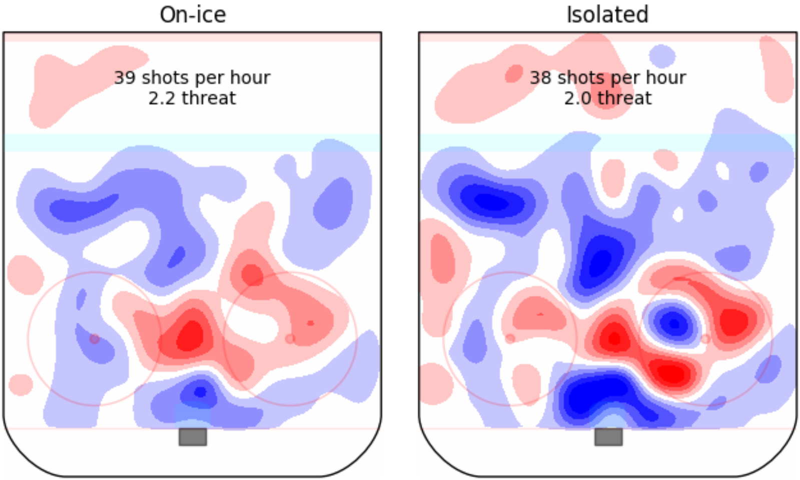


2017-2018 Henrik Sedin, Observed vs Isolated, Offence

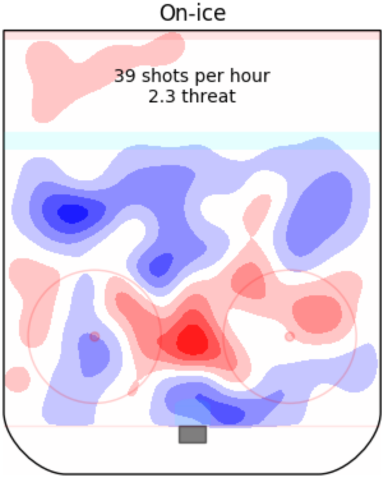


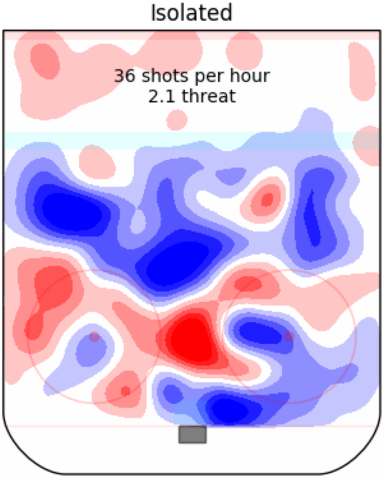


2017-2018 Daniel Sedin, Observed vs Isolated, Defence

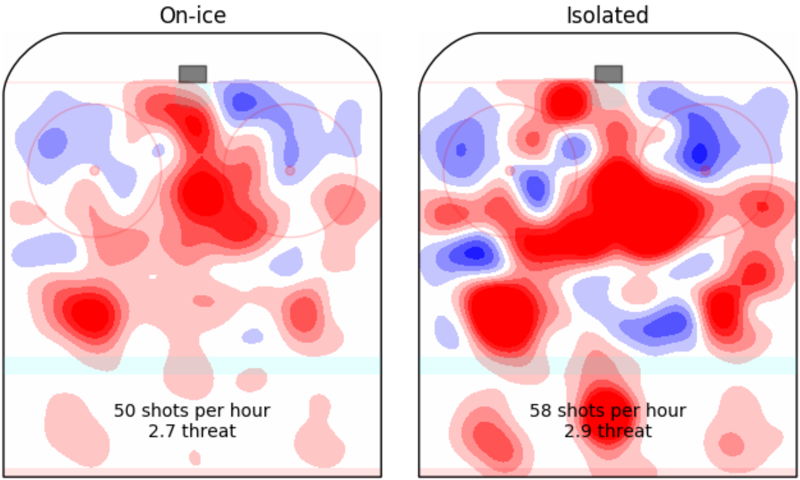


2017-2018 Henrik Sedin, Observed vs Isolated, Defence

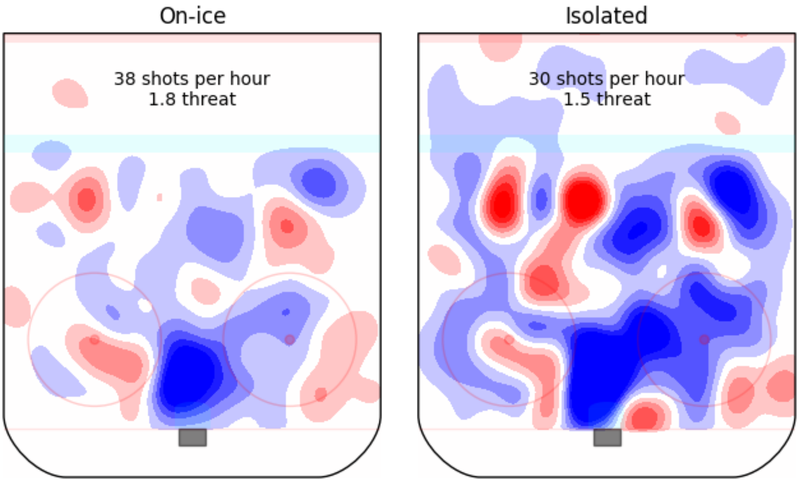




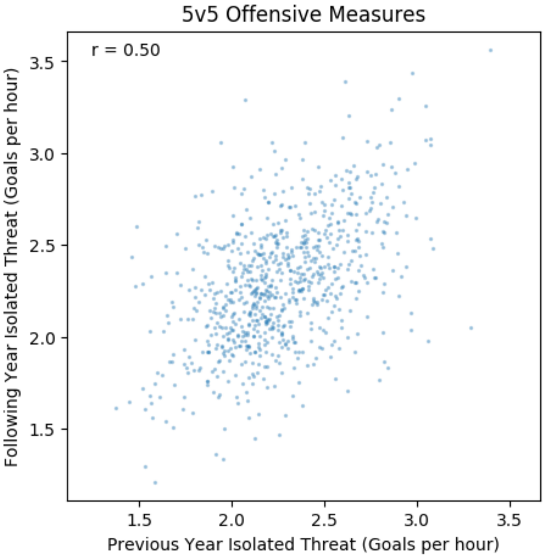
2017-2018 Mathieu Perreault, Observed vs Isolated, Offence



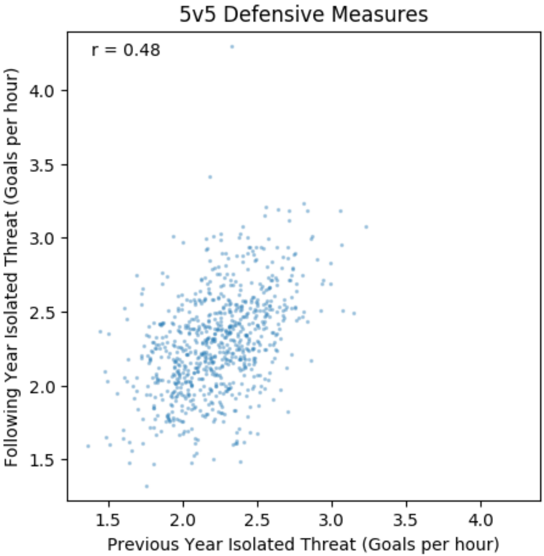
2017-2018 Mathieu Perreault, Observed vs Isolated, Defence



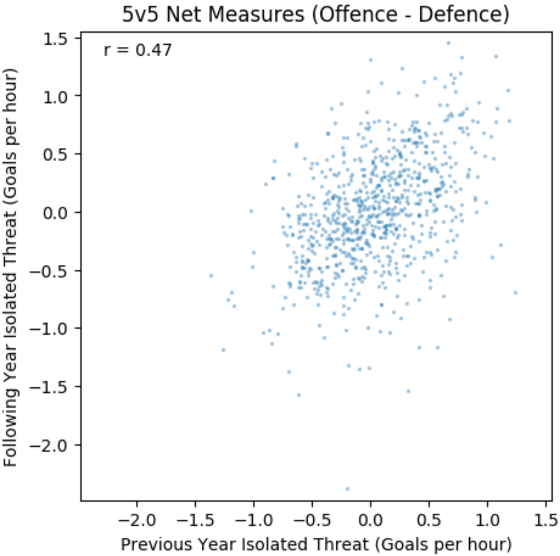
Correlations - Offensive Threat Created



Correlations - Defensive Threat Allowed



Correlations - Net Threat



Season-to-Season Auto-Correlations

For players who do not change teams, team-seasons for 2015-2016, 2016-2017, and 2017-2018, per hour of icetime:

	Offence	Defence	Net
On-ice Goals	0.36	0.14	0.18
On-ice Unblocked Shots	0.56	0.52	0.56
On-ice Shots	0.62	0.51	0.59
Isolated Threat	0.51	0.50	0.49

Season-to-Season Auto-Correlations

For players who do not change teams, team-seasons for 2015-2016, 2016-2017, and 2017-2018, per hour of icetime:

	Offence	Defence	Net
On-ice Goals	0.36	0.14	0.18
On-ice Unblocked Shots	0.56	0.52	0.56
On-ice Shots	0.62	0.51	0.59
Isolated Threat	0.51	0.50	0.49

Correlation between isolated offensive threat and isolated defensive threat per hour is 0.01. The two performances are independent.

Best Defensive Threat Performances, 2017-2018

Player	Team	Isolated Threat Against
Matt Stajan	CGY	1.32
Charles Hudon	MTL	1.38
Scott Laughton	PHI	1.45
Greg Pateryn	DAL	1.45
Mathieu Perreault	WPG	1.47
Colton Parayko	STL	1.48
Dmitrij Jaskin	STL	1.48
Jordan Nolan	BUF	1.52
Alexander Iafallo	LA	1.52
Dan Hamhuis	DAL	1.53

Best Offensive Threat Performances, 2017-2018

Player	Team	Isolated Threat For
Sidney Crosby	PIT	3.56
Timo Meier	SJ	3.44
Kris Letang	PIT	3.30
Conor McDavid	EDM	3.26
Michael Frolik	CGY	3.24
Joonas Donskoi	SJ	3.20
Pierre-Luc Dubois	CBJ	3.20
Patric Hörnqvist	PIT	3.08
Tyler Toffoli	LA	3.07
Markus Nutivaara	CBJ	3.06

Best Overall Threat Performances, 2017-2018

Player	Team	Isolated Threat Net
Mathieu Perreault	WPG	1.45
Hampus Lindholm	ANA	1.34
Kris Letang	PIT	1.33
Colton Parayko	STL	1.31
Mark Giordano	CGY	1.19
Joonas Donskoi	SJ	1.18
Pierre-Luc Dubois	CBJ	1.18
Dougie Hamilton	CGY	1.15
Timo Meier	SJ	1.14
Adam Lowry	WPG	1.12

Conclusions

- Isolated threat seems to describe repeatable aspects of
  - offensive,
  - defensive, and
  - all-around skater performance

  for players who do not change teams.

Future Work

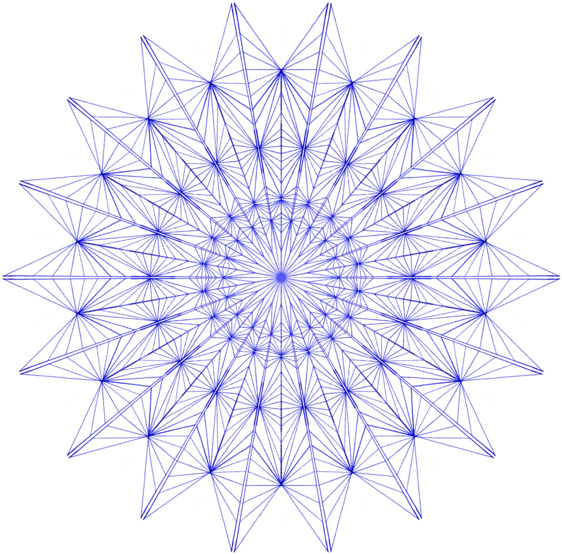
For shot map isolation itself:

- Quality-of-competition.
- Non-linear effects. (Chemistry!)

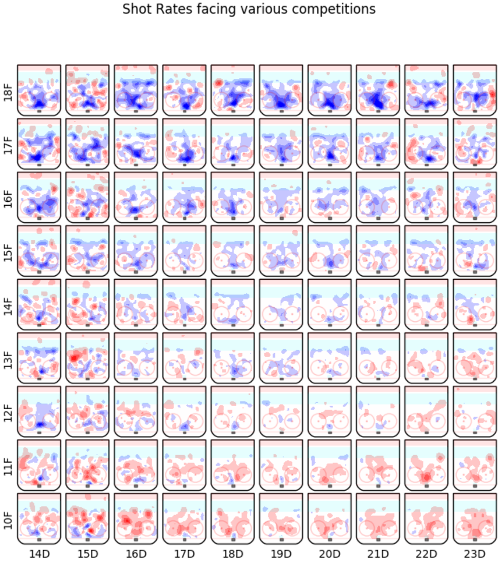
For a broader evaluation scheme:

- Goalies and shooting talent.
- Special Teams.

Thanks!



Quality of Competition



Quality of Competition