PUCK POSSESSED

VISUALIZING HOCKEY DATA AND ANALYSIS, FOR GENERAL FANS TO UNDERSTAND AND ENJOY

By @RJWEISE

Title: Puck Possessed

Subtitle: A visual introduction to hockey analytics for kids, adults (who have yet to grow up) and visual thinkers

Purpose of the book: Making hockey analytics available and accessible to kids, while explaining the basics and learning some basic math and statistics through reading.

Intended audience: Kids of early age (visually appealing) to junior high (math, stats), so K-9, adults, visual thinkers

For Lukas & Elliott. If they are the only two to read and like this book, it has been worth it.

This book would not be possible without the endless patience of my lovely wife Dolores.

++ Marc ++

Introduction and inspiration

There a many books on hockey on the market, of which a small number are about hockey analytics. And even fewer are written for an audience who may not yet be familiar with, or into the technical aspect of analytics, making for very theoretical (and visually unappealing) books that assume an understanding of many of these technical aspects. Actually, other than the Hockey Abstract and Stat Shot series by Rob Vollman, there is really not much else on the market for the hockey analytics newby. On the other hand there are again many kids' books on hockey with some being very visual and touching on analytics, but they typically just show yet not explain.

This book is written for these specific audiences:

- not yet familiar with hockey analytics;
- likes learning through visuals that are explanatory and understandable;
- Wants to be introduced to math and statistical concepts through a topic of interest.

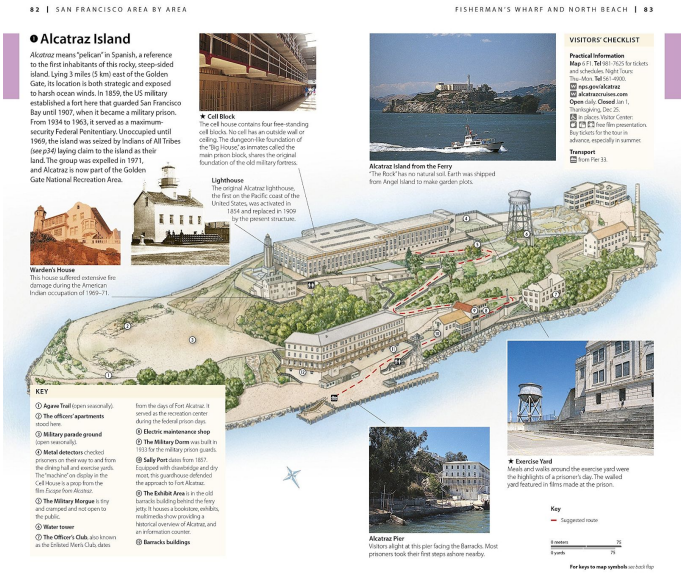
That being said, I still believe that the experienced hockey analytics audience can pick up a few concepts or ideas here and there that will increase their understanding of (the visualization of) hockey analysis. And for any reader I believe it will be a fun book to read.

Hi. I'm RJ. And I am puck possessed.

I grew up in The Netherlands, where football (or soccer) is the number one sport by a mile, and "hockey" is used to talk about field hockey rather than ice-hockey. And although I played the field type hockey for many years, for some reason the variant on ice always has appealed to me most. It wasn't until I was in my 20's that I finally started playing, but more at a level and seriousness that would best compare to shinny, or pick-up hockey. Except for my last year in The Netherlands, when I played inline-hockey (aka roller hockey) in the Northern Germany league, with a fun group of other Dutch Ice hockey fanatics. In short, I was never really able to play my most favorite kind of hockey

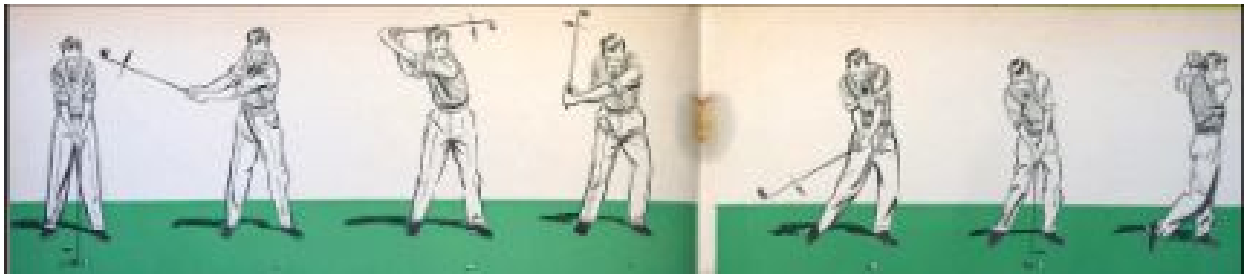
(it didn't help that The Netherlands has fewer hockey rinks than there are in Calgary alone!) and I am trying to catch up, even when my body is telling me not to.

In addition to a love for sport in general and hockey (from here I'm referring to the kind on ice) specifically, I have always loved illustrations, comics, and visuals explaining concepts, like these examples from DK books, described as a "visual feast" by a reviewer:



Not really believing in my own illustration capabilities, I studied to do spatial analysis and mapping in stead, which in time turned into data visualization of all kinds of data, not just spatia. And once I applied these skills to data about hockey, well, one thing leads to another and eventually you end up with a book like the one you are reading right now.

It is probably clear to you now that I love hockey, and that I am crazy about visuals that tell stories. Like the following, from a book about golf that I got at a book fair when I was still very young:



What I like a lot about the visual language is that it is universal. No matter what language you know (and don't know), the concepts of the golf swing (above) or from "Johan Cruyff geeft voetballes" (Johan Cruyff teaches soccer, right) are still understandable without understanding the actual written language through the visuals.

These type of books that I grew up with inspire me to write this hockey book in similar fashion. Not just because they are fun, but because they actually taught me about countries and buildings, how to swing a golf club and concepts about soccer. And they made me understand far better and deeper than what I am convinced textbooks would have been able to achieve.

Over the years I have ran into quite a few of these visual explanatory and exploratory books, most of them related to sports one way or another. Some of these are shown a little further below, with examples of visuals being used to explain different concepts in sports: strategy, X's and O's, technique, positioning, uniforms, line-ups, and more. They really gave me many "aha" moments, all of them inspired me then to learn more, and now, to write this book.

The content of the book is both “interesting” and instructional; it starts with looking at the different types of “analysts” in relation to hockey or other sports, followed by some concepts around the required elements of data analysis and visualization: data, tools, math & statistics, analytics, visualization and practical applications of all the learned skills. It will also include examples of math and statistical concepts based on the Kindergarten to grade nine level of math to allow younger readers to go along and pick up some skills they can apply in the classroom.

I hope the readers of this book will have as much fun and enjoyment as I had, and better understand hockey analytics in general and how it is applied today. Enjoy!

How “nerdy” do you need to be?

This book is focussed on adults new to hockey analytics and kids and their parents who like hockey first and foremost, and who like to play with statistics beyond points and +/- to better understand the quality and value of players. Some refer to those people as *nerds*. And while the reasoning for that stigma escapes me, fortunately we live in a time where that is actually a good thing.

Do you like to get lots of statistics and play with the data? Or do you like to focus on one topic, get specific related data and analyze it to find some hidden secrets? Perhaps you are not too much in to data crunching but like to focus on visualizing the data so that it highlights your findings and communicates that clearly to people that read your blog? Maybe you are just a fantasy hockey player who wants to get the upper hand by better understanding hockey data than your opponents, or perhaps you just want to understand your team better in the players they trade for, and if your team plays at their level or if they are (un)lucky. And then there is the know it all, who just wants to know as much as he or she can to be the one to answer everyone’s trivia about hockey.

Most of you may be a little bit of this and a little of that, and some may not relate to any of these description at all, or strongly. It doesn’t really matter. This book is for all of you. If you like numbers but understand that they don’t replace expert eyes when evaluating players or teams but are a different view in addition to the expert’s eye, than we’re going to have fun.

So let start with the basics: what are data or statistics, and how do we get them? According to Wikipedia, ***Statistics*** *is a branch of mathematics dealing with data collection, organization, analysis, interpretation and presentation of masses of numerical data (Merriam-Webster dictionary). While many scientific investigations make use of data, statistics is concerned with the use of data in the context of uncertainty and decision making in the face of uncertainty.*

Wow, that’s a mouth full. Let’s simplify that a bit:

- We're going to be dealing with numbers a lot so that somehow some math gets involved is no big surprise;
- We will need stats, or data, so we need to collect it in some way or another;
- To work with the data properly we need to organize it, which includes cleaning of data and putting it in a format that works for our purpose and tool(s) of choice;
- We want to know what this data means, and tells about hockey players and teams, so we do some analysis;
- We need to know what we find means, and what it specifically means in our context of hockey;
- Lastly, unless we want to keep it all to ourselves, we want to present our data in a way to communicates the meaning of it based on the results of our analysis.
- (oh right, and we'll probably be dealing with lots and lots of data, although that is a fairly relative term.)

I think in terms of hockey analytics, data are often referred to when people talk about statistics. So what is data then, one might ask. Let's go back to Wikipedia:*Data is a set of values of subjects with respect to qualitative or quantitative variables. Data and information or knowledge are often used interchangeably; however data becomes information when it is viewed in context or in post-analysis. Data is measured, collected and reported, and analyzed, whereupon it can be visualized using graphs, images or other analysis tools. Raw data ("unprocessed data") is a collection of numbers or characters before it has been "cleaned" and corrected by researchers.*

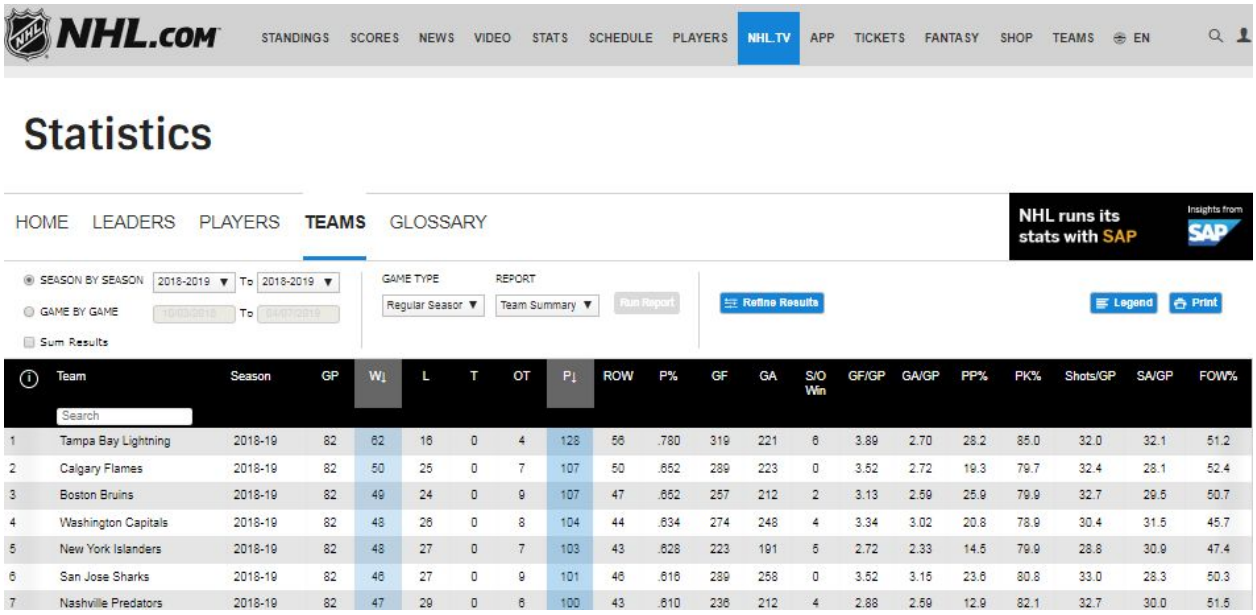
Ok. There appears to be quite the overlap between data and statistics, at least in the sense of how we talk about them. I'm going to argue here that we mostly need data, and often raw data, to which we may apply some statistical analysis.

Let’s start with the basics: data

There are a number of sources that we can use, of which NHL.com is the most obvious. Unfortunately the site does not let you easily grab or download the data in bulk, but we’ll get to that later on. For now we’ll just assume you can copy & paste the data we will be talking about. It has a lot of good data, under the tab called *Stats* (did we not just talk about that?). They provide leaderboards, player and team stat, and also provide a glossary explaining what everything means, something not available on every website with hockey data. Hockey-reference.com is another great site to get data, and it does allow for downloading the files to your machine. It also provides the Player Index, a tool to filter specific data rather than a (number of) season(s).

Both NHL and Hockey-reference provide aggregated or summarized data, so rather than getting data shot-by-shot, you’ll see something like player X had 5 shots in this game, or 123 shots this season. Granularity, or level of detail of the data is important to think about when you define the project you want to work on. You can imagine for a historic trend in number of goals in the NHL a season by season total would be enough. If you want to do the project for a specific team you’d have to go a bit deeper, getting shots per season per team. For players you have to go even further, getting shots per season, per team, per player. When you also remember that players get traded from one team to another, sometimes mid-season, an original simple idea becomes more complicated fast.

Besides these two sources there are a number of websites managed by hockey fanatics who collect, clean and analyse some of this data and offer it for download. Some for free, others for a subscription fee. This varies from high level, seasonal and team-level advanced statistics to highly detailed where every event in a game is tracked (think, shots, hits, goals, saves, etc.). To name a few: NHL, Natural Stat trick, Evolving Wild, and Corsica.



NHL.com STANDINGS SCORES NEWS VIDEO STATS SCHEDULE PLAYERS NHL.TV APP TICKETS FANTASY SHOP TEAMS EN

Statistics

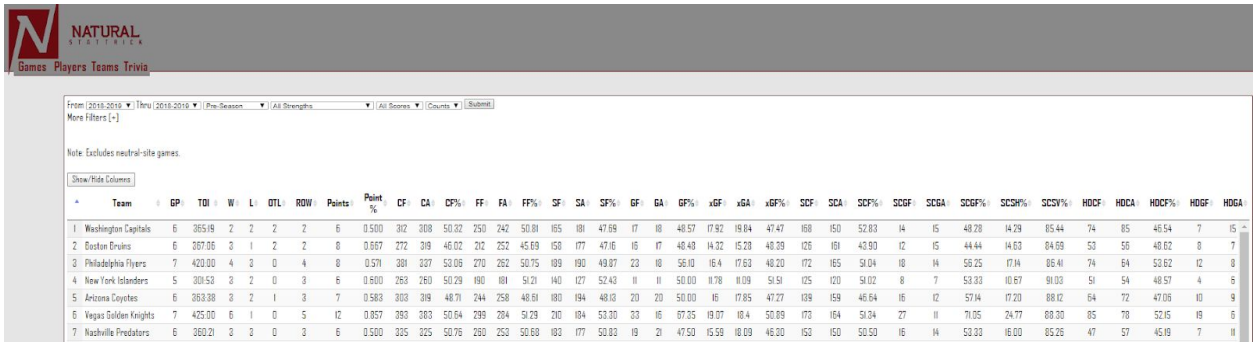
HOME LEADERS PLAYERS TEAMS GLOSSARY

NHL runs its stats with SAP — Insights from SAP

SEASON BY SEASON 2018-2019 To 2018-2019 | GAME BY GAME | Sum Results

GAME TYPE: Regular Season | REPORT: Team Summary | Run Report | Refine Results | Legend | Print

	Team	Season	GP	W	L	T	OT	P	ROW	P%	GF	GA	S/O Win	GF/GP	GA/GP	PP%	PK%	Shots/GP	SA/GP	FOW%
	Search																			
1	Tampa Bay Lightning	2018-19	82	62	16	0	4	128	56	.780	319	221	6	3.89	2.70	28.2	85.0	32.0	32.1	51.2
2	Calgary Flames	2018-19	82	50	25	0	7	107	50	.652	289	223	0	3.52	2.72	19.3	79.7	32.4	28.1	52.4
3	Boston Bruins	2018-19	82	49	24	0	9	107	47	.652	257	212	2	3.13	2.59	25.9	79.9	32.7	29.5	50.7
4	Washington Capitals	2018-19	82	48	26	0	8	104	44	.634	274	248	4	3.34	3.02	20.8	78.9	30.4	31.5	45.7
5	New York Islanders	2018-19	82	48	27	0	7	103	43	.628	223	191	5	2.72	2.33	14.5	79.9	28.8	30.9	47.4
6	San Jose Sharks	2018-19	82	46	27	0	9	101	46	.616	289	258	0	3.52	3.15	23.6	80.8	33.0	28.3	50.3
7	Nashville Predators	2018-19	82	47	29	0	6	100	43	.610	236	212	4	2.88	2.59	12.9	82.1	32.7	30.0	51.5



NATURAL STATTRICK

Games Players Teams Trivia

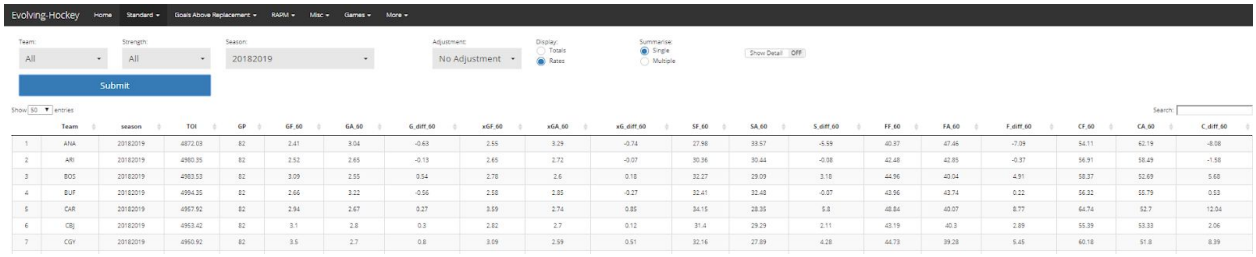
From 2018-2019 Thru 2018-2019 Pre-Season All Strengths All Scores Counts Submit

More Filters [+]

Note: Excludes neutral-site games.

Show/Hide Columns

	Team	GP	TOI	W	L	OTL	ROW	Points	Point %	CF	CA	CF%	FF	FA	FF%	SF	SA	SF%	GF	GA	GF%	xGF	xGA	xGF%	SCF	SCA	SCF%	SCGF	SCGA	SCGF%	SCSH%	SCSV%	HDCF	HDCA	HDCF%	HDGF	HDGA
1	Washington Capitals	6	365:19	2	2	2	2	6	0.500	312	308	50.32	250	242	50.81	165	181	47.69	17	18	48.57	17.92	19.04	47.47	168	150	52.83	14	15	48.28	14.29	85.44	74	85	46.54	7	15
2	Boston Bruins	6	367:06	3	1	2	2	8	0.667	272	319	46.02	212	252	45.69	158	177	47.16	16	17	48.48	14.32	15.28	48.39	126	161	43.90	12	15	44.44	14.63	84.69	53	56	48.62	8	7
3	Philadelphia Flyers	7	420:00	4	3	0	4	8	0.571	381	327	53.06	270	262	50.75	199	190	49.87	23	18	56.10	16.4	17.63	48.20	172	165	51.04	18	14	56.25	17.14	86.41	74	64	53.62	12	8
4	New York Islanders	5	301:53	3	2	0	3	6	0.500	263	260	50.29	190	181	51.21	140	127	52.43	11	11	50.00	11.78	11.09	51.51	125	120	51.02	8	7	53.33	10.67	91.03	51	54	48.57	4	6
5	Arizona Coyotes	6	363:38	3	2	1	3	7	0.583	303	319	48.71	244	258	48.61	180	194	48.13	20	20	50.00	16	17.85	47.27	139	159	46.64	16	12	57.14	17.20	88.12	64	72	47.06	10	9
6	Vegas Golden Knights	7	425:00	6	1	0	5	12	0.857	393	303	50.64	299	284	51.29	210	184	53.30	33	16	67.25	19.07	18.4	50.89	173	164	51.34	27	11	71.05	24.77	88.30	85	78	52.15	18	6
7	Nashville Predators	6	360:21	3	3	0	3	6	0.500	335	325	50.76	260	253	50.68	183	177	50.83	19	21	47.50	15.59	18.09	46.30	153	150	50.50	16	14	53.33	16.00	85.26	47	57	45.19	7	11



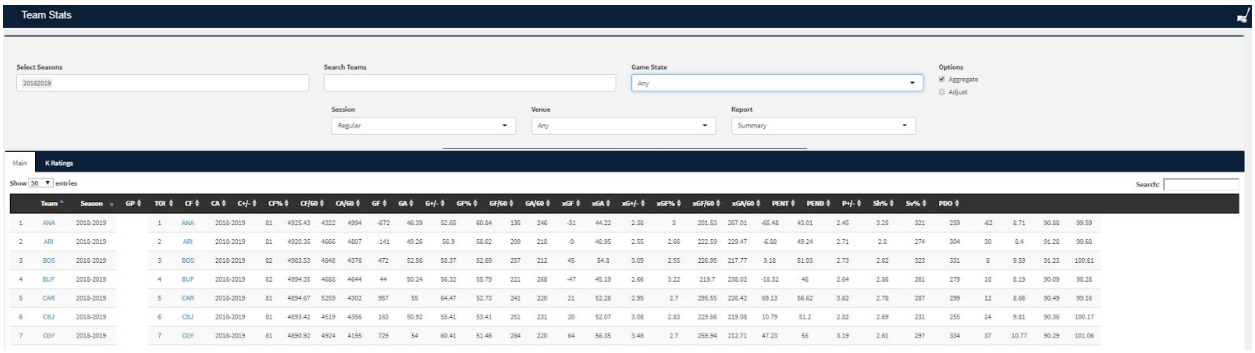
Evolving-Hockey Home Standard Goals Above Replacement RAPM Misc Games More

Team: All | Strength: All | Season: 20182019 | Adjustment: No Adjustment | Display: Totals / Rates | Summarize: Single / Multiple | Show Detail: Off

Submit

Show 50 entries Search:

	Team	season	TOI	GP	GF_60	GA_60	G_diff_60	xGF_60	xGA_60	xG_diff_60	SF_60	SA_60	S_diff_60	FF_60	FA_60	F_diff_60	CF_60	CA_60	C_diff_60
1	ANA	20182019	4872.03	82	2.41	3.04	-0.63	2.55	3.29	-0.74	27.98	33.57	-5.59	40.87	47.46	-7.09	54.11	62.19	-8.08
2	ARI	20182019	4980.35	82	2.52	2.65	-0.13	2.65	2.72	-0.07	30.36	30.44	-0.08	42.48	42.85	-0.37	56.91	58.49	-1.58
3	BOS	20182019	4983.53	82	3.09	2.55	0.54	2.78	2.6	0.18	32.27	29.09	3.18	44.96	40.04	4.91	58.37	52.69	5.68
4	BUF	20182019	4994.35	82	2.66	3.22	-0.56	2.58	2.85	-0.27	32.41	32.48	-0.07	43.96	43.74	0.22	56.32	56.79	0.53
5	CAR	20182019	4967.92	82	2.94	2.67	0.27	3.59	2.74	0.85	34.15	28.35	5.8	48.84	40.07	8.77	64.74	52.7	12.04
6	CBJ	20182019	4963.42	82	3.1	2.8	0.3	2.82	2.7	0.12	31.4	29.29	2.11	43.19	40.3	2.89	55.39	53.33	2.06
7	COY	20182019	4960.92	82	3.5	2.7	0.8	3.09	2.59	0.51	32.16	27.89	4.28	44.73	39.28	5.45	60.18	51.8	8.39



So now we know where to find some data, but how do we get it into our (virtual) hands?

Based on your interests your preferred way of collecting the data from websites will be one of the following: copy & pasting, writing code, or paying a subscription or one -time payment. There are some alternatives somewhere in between, but before we dig deeper into these alternatives, let's talk about live vs. static data. If you want to show the number of shots taken per team, per season, you can download one big data set for all previous completed seasons, and add the current season when it is concluded. Those datasets will not change over time, so one download will be all you need. But how about measuring a shooters shooting percentage as the season progresses? Or how the standings change from day to day? In that case we can still load data manually every day and add it to our existing data set, but that becomes quite a bit of (boring) work. This is where "live" data is more useful. With "live" data I mean data that get updated on a daily or gamy-by-game basis. Let's start with static, or one-time data.

We can go to the website of our choice and load the table with the data we want, select all the rows and columns of the table, copy the data and then paste it in Excel, Google Sheets or something similar. Unfortunately not every website works that simple anymore and we may have to learn some HTML (the language used to make websites) to specifically identify the table on a webpage that we want to download. Copy & paste works fine for small datasets, say 31 records (one per team in the nhl), but even when we want data for all players, typically one table page is not sufficient. Then we need to copy-paste the data from the first table-page, load page 2, copy & paste, load page 3, etc. Quite the effort if there are 10 pages or more! There must be a better way!

And there often is. Google sheets have functions to load html tables either directly or through specific XML-paths, and there are some scraper tools available that claim they can find and download the table's full dataset. People that like to (learn to) write code can use languages like Python and R to write a few lines to accomplish the same, while storing the collected data in one complete dataset.

Some other websites offer (often small) subscription fees, basically to keep the computers and servers running that make the data available. Subscribers can go to these sites, log in, and download the data whenever they need in an easy format like Excel or CSV.

Some websites offer what they call API's, backdoor entrances to the sites data, often raw, that can then be collected and used. Wikipedia describes and API as follows: *An Application Programming Interface (API) is a set of functions, procedures, methods or classes used by computer programs to request services from the operating system, software libraries or any other service providers running on the computer.*

To do so one often needs to pay some money, as well as have some knowledge of how to make this all work. These APIs, and the NHL provides a very popular one that has pretty much all data they collect, require technical coding (or programming) skills to collect, clean and format the data, and then export the data to be used for analysis and / or visualization. Other than providing a lot of data, another benefit of these API's is that they are typically up-to-date and continuously refreshed when new data becomes available, often even during a game! Some of the previously mentioned subscription services actually get their data from the NHL API, clean and format it, and then make it easily available to their subscribers.

Another thing to consider is what data to download. Most services and websites offer the option to for example download events for all strengths, only 5 on 5, powerplay and shorthanded, etc. For player data they often give the option to only use data from players that have a played a certain number of games, or data from only the regular season, since playoff hockey is so different.

Once we have the data we want, typically we find out that there are some issues with it, so we'll need to "clean it up". Column names like +/-, slight differences in playernames and team abbreviations, Time on ice on minutes, seconds or hours:minutes:seconds,more than one player with the same name, different position descriptions, etc. Despite these often being small issues, cleaning data can - and often does - take a significant amount of time.

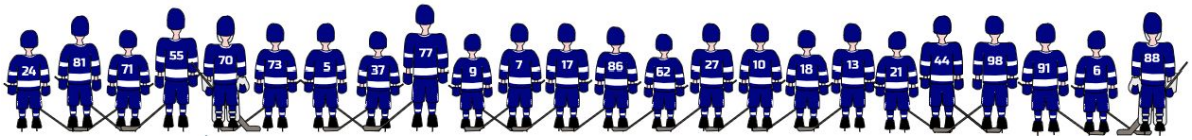
To avoid having a lot of data and only use part of it - which would not be great for performance and make it harder to work with - , we'll want to remove the data that we don't need. Ideally we do this by keeping the main dataset as it is and create a subset that meets our specific needs. In that case we can always go back one step rather than having to start all over.

Now that we have a better idea of what we need to do to collect and clean data, let's talk about math.

Math & Statistics (RJ: use K-3, 4-6, 7+)

Learning numbers and solving problems with patterns (K)

To learn about numbers, we're going to start with the best team of the regular season, the Tampa Bay Lightning:



What numbers do you recognize? And what is your favorite number? Who do you think the defencemen are, the smaller or the taller players?

1 - Building the basics through addition and subtraction

Other than what numbers are and counting, the first thing we learn to do with numbers is recognizing patterns, adding and subtracting. Many statistics are counts of specific events; things like how many players are on the ice in total or per team, do both teams have the same number of players, how many teams are in each Division of the Eastern & Western conference, and do they have the same number of teams?

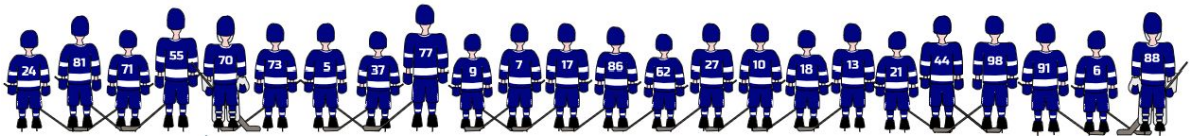
Metropolitan Division	Pacific Division	Atlantic Division	Central Division
Team	Team	Team	Team
Capitals	Flames	Lightning	Predators
Islanders	Sharks	Bruins	Jets
Penguins	Golden Knights	Maple Leafs	Blues
Hurricanes	Coyotes	Canadiens	Stars
Blue Jackets	Canucks	Panthers	Avalanche
Flyers	Ducks	Sabres	Blackhawks
Rangers	Oilers	Red Wings	Wild
Devils	Kings	Senators	

To get a quick idea of the answer to some of these questions we can subitize (recognize at a glance) and see familiar

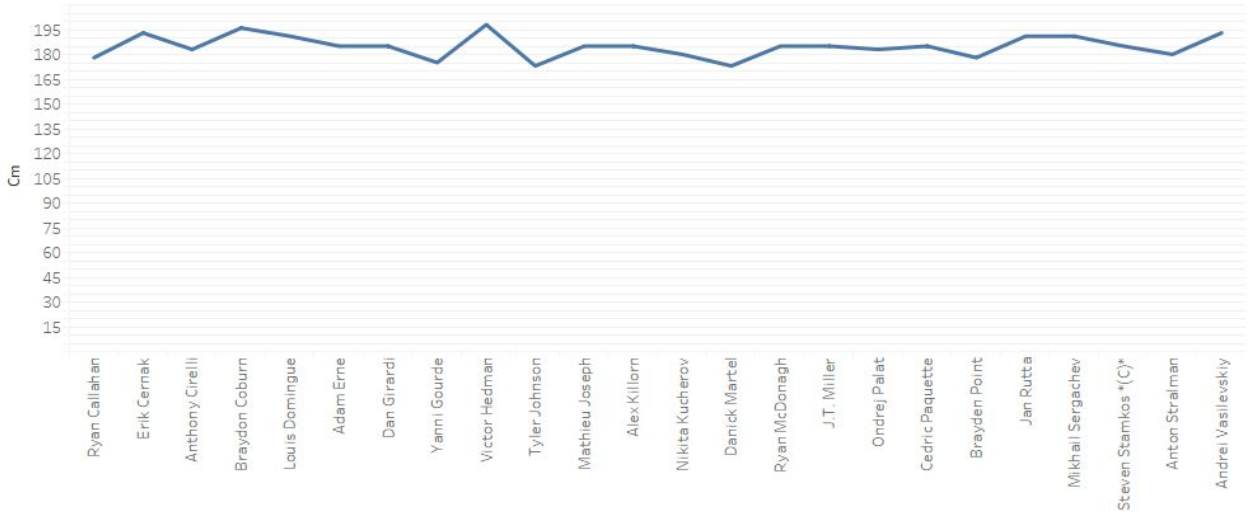
arrangements. All the questions above can be answered by adding and subtracting players and / or teams.

2 - Using mathematics to solve problems.

As we get more comfortable with counting, adding and subtracting, we will count, describe and estimate quantities in a variety of ways. We will solve problems using numbers, patterns, measurement and data collection, and use graphs and charts to communicate information.



If we put the players from above in a chart based on their heights, they would look like this:



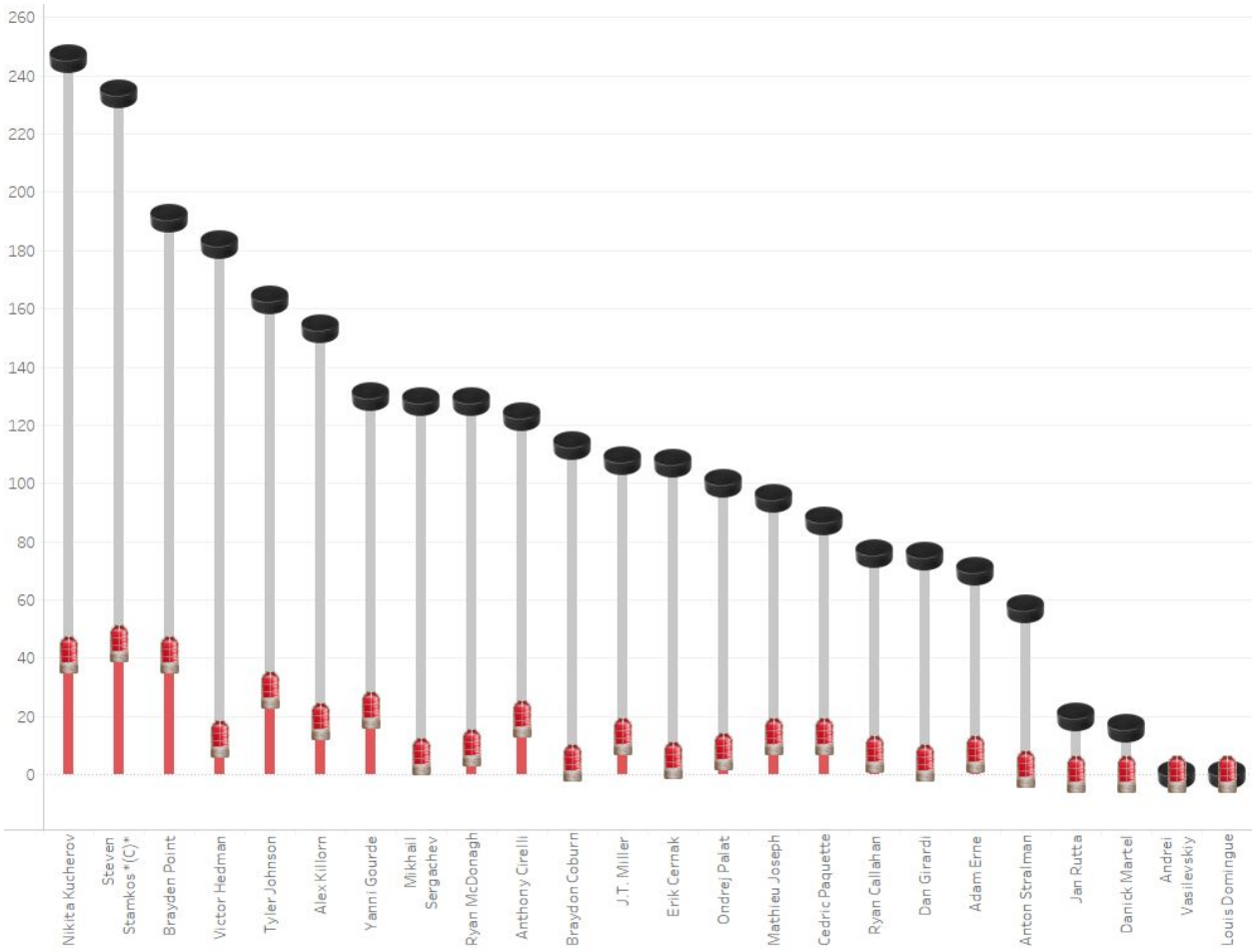
So far we ordered the players alphabetically by their last name. Now let's order them based on their height:

The two goalies are amongst the tallest players on the team, and Adam Erne (#73) is the tallest forward but seven players are taller than him. Anton Stralman (#6) is the smallest defenceman but there are still 6 players smaller than he is.

3 - Introduction to multiplication and division.

From here we will understand, apply and recall addition facts and related subtraction facts, and use mental mathematics strategies. Multiplication and division are the new topics we'll add to our skill set.

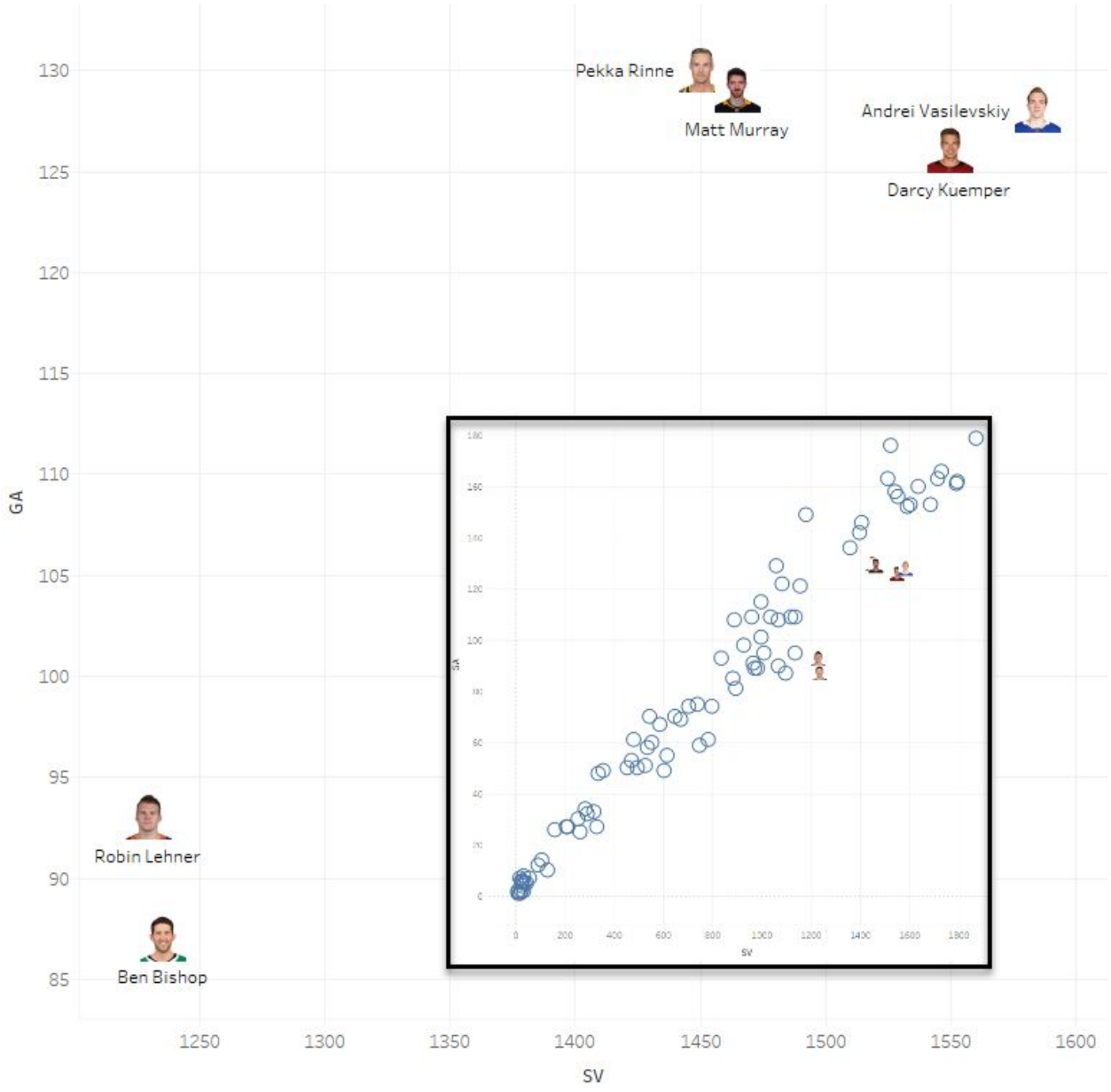
So, the simplest form of math applied when collecting statistics, or data, is counting stats. How many shots did player X take in a game, how many power play opportunities did team Y have, how often did goalie Z make a save, etc. We can simply collect these by counting the events we are interested in, so all you need to be able to do is count. And perhaps some addition, multiplication and division to aggregate or summarize the data for a full season.



4 - Learning about fractions and decimals.

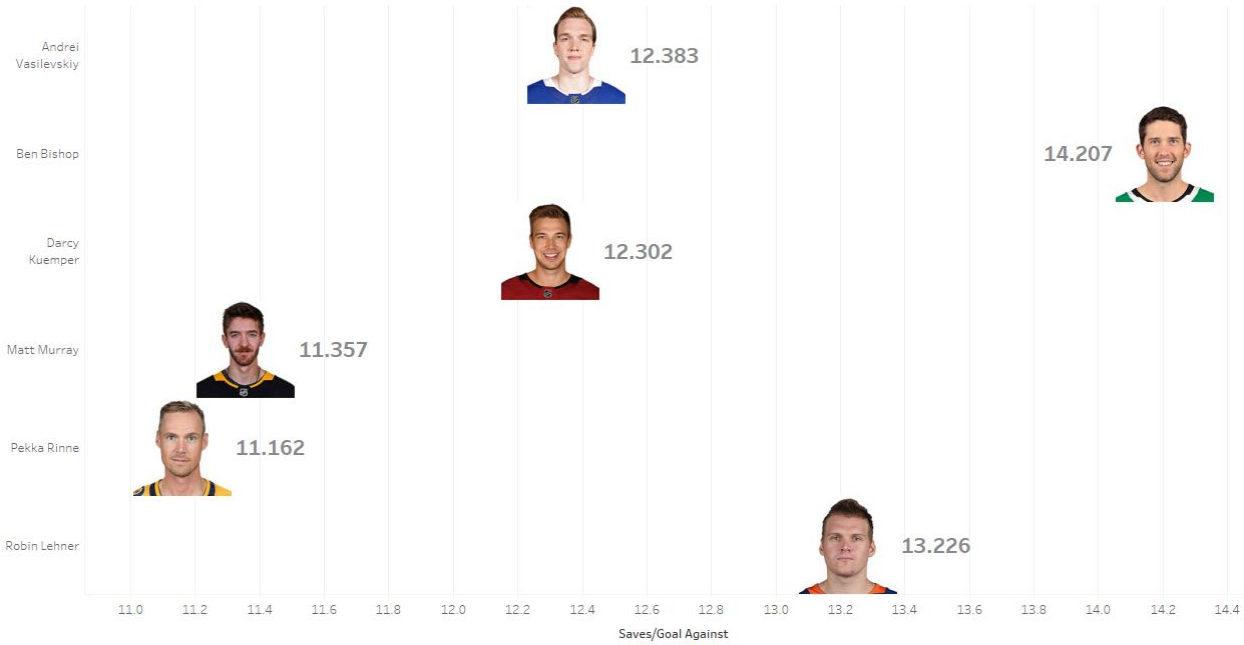
Now that we start multiplying and dividing, we will run into decimals and fractions. <...> which are?

In Puck Possessed issue 16 I talk about ranking goalies by number of saves per goal against. More specifically, how many shots does a goalie stop before he let's one go in? The following chart shows some of the best goalies from the 2018-2019 regular season when looking at Goals Against and Saves made:



The insert shows all goalies and where the highlighted goalies sit relative to the league.

When we divide Saves made by Goal Against here is what they look like:



So let's look at Ben Bishop from the Dallas Stars: during the regular season he had 87 goals against while making 1,236 saves. When we divide 1,236 / 87 we see that he made a little over 14 saves per goal against. In the chart above we use three decimals, but you can ask if that really provides useful information. I'd say one decimal would be sufficient here, so 14.2 saves per goal against (14.1 or 14.9 is still worth mentioning, but 14.17 or 14.95 doesn't provide a better understanding).

5 - Using mathematics to solve problems.

Ok, we understand, recall and can apply multiplication and related division and use mental mathematics strategies. We can compare fractions with like and unlike denominators, and describe, compare, add and subtract decimal numbers and lastly we learn about probability (although here we'll leave that out until we get to part 7).

To apply all this to hockey, we can compare players and decide who is the best scorer; but just counting goals is not really fair. One player will have other opportunities than

other players, based on the circumstances the player is in, of which many we cannot control or even measure easily; is the team he plays on very good or bad, is his coached offensive or defensive, how much ice time does he have, who does he play with, who is the opposing goaltender, etc. What we can do is use rates or ratios, for which we will have to do some division. Rather than only looking at the number of goals, let's look at the number of goals per games played. So let's divide number of goals by games played:

G / GP

This gives us a much better idea of who is the better scorer. Is it fair to not take into consideration if a player has been sick or injured and hasn't played much, compared to a player who has played every single game? I would say it is not. We can even take this a bit further and look at goals per minutes played, making the math a bit more complicated but the results even more useful.

6 - Operations with numbers.

So ratios are not much more than one counting stat per some other counting stat; goals per minute played, saves per goal allowed. When we standardize this to 100 minutes or goals allowed or look at counting stat of a player per counting stats of all players (percent of whole), we use percents (which literally means "per 100"). Also we need to start thinking about the order of operations. The rules are as follows:

1. Do operations inside parentheses.
2. Do multiplication and division from left to right.
3. Do addition and subtraction from left to right.

Another way of using ratios is to look at goals per shot attempt, or goals per 60 or 20 minutes. Typically people use per 60, as a game lasts 60 minutes, but I believe a rate per 20 makes a lot of sense as a players time on ice is more in the 20 minute range, or perhaps even 15. In any case, one would look at the number of goals compared to the time the player was on the ice, and since no player have has the exact same time on ice as any other player we convert it to 15/20/60 minutes:

G / TOI * 60

So what if we want to know the shooting percentage for a whole team? We use the average of all players on that team, or to avoid using numbers from players that only played a few games (which is not a good representation of a player's skills. So let's say on a full season of 82 games, a player needs to play at least 20 games to be included in the calculation of the team average:

For all players that have a GP of 20 or more: **SUM(G) / SUM(GP)**

The results of calculations like these often have a large number of decimals. Although we understand that if you look at the average number of goals by players in the same team will not always be a full number, having 10 decimals also doesn't add any value. So let's say for now we use one decimal.

7 - Learning about statistics.

As mentioned above we are now looking at averages, the mean, the median and the mode for a set of data; percents, rates, ratios and proportions, and lastly correlation and probability. Let's look at the Lightning team again, and the players' ages:The mean is calculated by adding up all of the values and dividing by the number of values.

Player	Age
Ryan Callahan	33
Erik Cernak	21
Anthony Cirelli	21
Braydon Coburn	33
Louis Domingue	26
Adam Erne	23
Dan Girardi	34
Yanni Gourde	27
Victor Hedman	28
Tyler Johnson	28
Mathieu Joseph	21
Alex Killorn	29
Nikita Kucherov	25
Danick Martel	24
Ryan McDonagh	29
J.T. Miller	25
Ondrej Palat	27
Cedric Paquette	25
Brayden Point	22
Jan Rutta	28
Mikhail Sergachev	20
Steven Stamkos *(C)*	28
Anton Stralman	32
Andrei Vasilevskiy	24

The **mean** is calculated by adding up all of the values and dividing by the number of values:

Sum of ages: 633, and number of players: 24. The mean, or average age = 633 / 24 = 26.375

The **median** the "middle" of a set of numbers in ascending or descending order: 26.5. Since there are 24 ages, when ordered from low to high the middle is where there are 12 ages to the left and 12 to the right. These are 26 and 27 > middle is 26.5*

The mode is the most frequently occurring number:28, as there are 4 players that are 28 years old, more than any other age.

* Median example:

20 21 21 21 22 23 24 24 25 25 25 26 27 27 28 28 28 28 29 29 32 33 33 34

We already talked about percents, ratios and proportions above but let's repeat to refresh our memory:

- Percent means "out of one-hundred", so if a goalie (remember Bishop?) saves 1,236 shots and let's 87 puck go in the net, we say he has a 0.9296 save percentage, or in other words he saves 93 of every 100 shots saved (93 per cent).
- A ratio is a comparison of numbers or quantities, dividing one by the other. Bishop's goals ratio compared to shots faced is 87 / 1,236 is 0.07.
- A proportion is a statement of equality between two ratios.

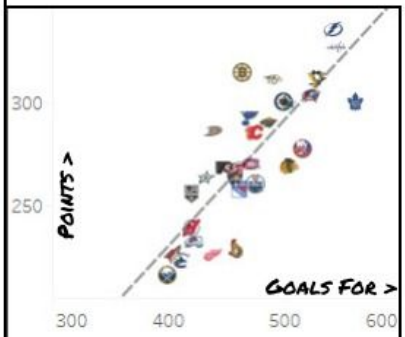
Now let's talk about correlation. As quoted in Puck Possessed issue 21 Wikipedia states correlation as *any statistical association, though it commonly refers to the degree to which a pair of variables are linearly related...* Correlations are useful because they can indicate a predictive relationship that can be exploited in practice... However, in general, the presence of a correlation is not sufficient to infer the presence of a causal relationship (i.e., correlation does not imply causation)... There are several correlation coefficients, often denoted, measuring the degree of correlation.

The mentioned issue 21 is included below to give some example to better explain the concepts:

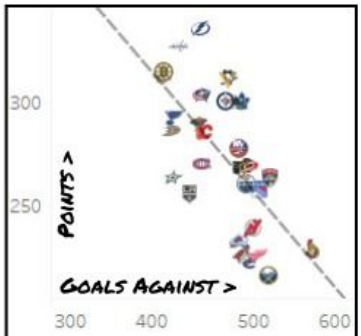
PUCK POSSESSED #21

CORRELATION

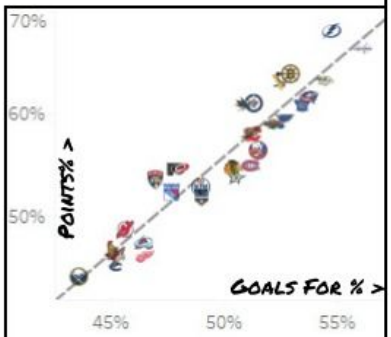
Wikipedia describes correlation as follows: ...any statistical association, though it commonly refers to the degree to which a pair of variables are linearly related ... they can indicate a predictive relationship. In this issue we investigate what stats can predict team points. Let's start with something simple: goals for and against related to points over the 2016-17, 2017-18 and2018-19 regular seasons, 5 on 5 data



We can see the relationship, or correlation here between goals for and points; the more goals a team scores, the more points they have. Although not perfect (then they would be on the diagonal line) there is a strong correlation, expressed in R-square. In this case R-square is about 67%.

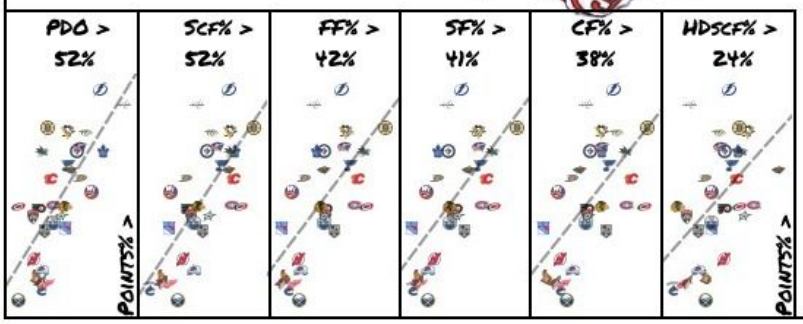


With goals against, the relation is negative; the more goals against, the lower the points. But the correlation is still strong, at about 41%



If we look at point% (points reached compared to total possible points) and goals for (goals for divided by all goals (for and against)), all teams are close to the diagonal, hinting a strong correlation. R-sqaure is almost 90%.

What about other statistics? Do they have a better correlation, and thus predicitve value for team points? In the charts below we look at a few (>), compared to points % (^)



Based on the analysis in this issue the Goals For % has the highest correlation with Points%, which suggests we use these stats when trying to predict team points based on player projections for upcoming seasons. Pretty handy for you fantasy hockey leagues!

IMAGE SOURCES

Player pile up picture from Chris Lee: clee@post-dispatch.com

DATA Natural Stat Trick; see their glossary for descriptions of the used abbreviations

Connect on Twitter: @RJWEISE

Brief overview of Hockey Analytics & Visualization

There are a number of books and websites that provide a thorough overview on the history of hockey analytics, so it wouldn't really make sense to do the same thing all over again. The more important questions is: What is hockey analytics, and why do we want to use it?

Wikipedia ("analytics"): Analytics is the discovery, interpretation, and communication of meaningful patterns in data; and the process of applying those patterns towards effective decision making. In other words, analytics can be understood as the connective tissue between data and effective decision making, within an organization.

Wikipedia ("hockey analytics"): the analysis of the characteristics of hockey players and teams through the use of statistics and other tools to gain a greater understanding of the effects of their performance. Three commonly used statistics in ice hockey analytics are "Corsi" and "Fenwick", both of which use shot attempts to approximate puck possession, and "PDO", which is often considered a measure of luck.

Understanding what you are trying to show - basic terminology (corsi is really just rate of puck possession through shot attempts)

If we go back to the Wikipedia definition of Analytics, the visualization part is described as "communication of meaningful patterns in data". So once we are done our discovery and interpretation of the data, and have drawn a conclusion with regards to the topic of choice, we need to communicate this to the intended audience. Again, this can be done a different levels of detail: league, team, player, events.

Knowing your audience is very important, as it determines how you communicate, at what level of detail and in what format. Is the audience knowledgeable, experienced, young or old, able to spend time or in need of a quick visual, able to interact with the visual (computer) or not (print), etc.?

A visualization needs a balance between effective communication (Tufte's ink ratio) and design (to attract) to make it attractive, clear, actionable and memorable.

Wikipedia ("**infographic**"): Infographics (a clipped compound of "information" and "graphics") are graphic visual representations of information, data or knowledge intended to present information quickly and clearly.[1][2] They can improve cognition by utilizing graphics to enhance the human visual system's ability to see patterns and trends.

Wikipedia ("**Data visualization**"): Data visualization is viewed by many disciplines as a modern equivalent of visual communication. It involves the creation and study of the visual representation of data.[1]

To communicate information clearly and efficiently, data visualization uses statistical graphics, plots, information graphics and other tools. Numerical data may be encoded using dots, lines, or bars, to visually communicate a quantitative message.[2] Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable. Users may have particular analytical tasks, such as making comparisons or understanding causality, and the design principle of the graphic (i.e., showing comparisons or showing causality) follows the task. Tables are generally used where users will look up a specific measurement, while charts of various types are used to show patterns or relationships in the data for one or more variables.

Data visualization is both an art and a science.

Wikipedia ("**Information visualization**"): Information visualization or information visualisation is the study of (interactive) visual representations of abstract data to reinforce human cognition. The abstract data include both numerical and non-numerical data, such as text and geographic information. However, information visualization differs from scientific visualization: "it's infovis [information visualization] when the spatial representation is chosen, and it's scivis [scientific visualization] when the spatial representation is given".[1]

Putting it all to use

Fantasy

Publications

Tableau Public

Google Sheets

Extracting data

=IMPORTHTML

-NHL data

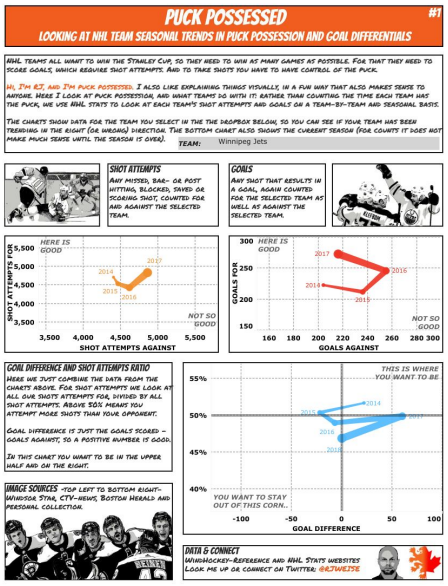
-R scripts

Services

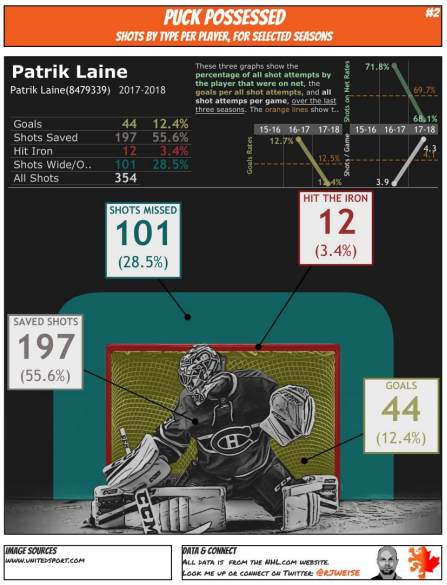
Puck Possessed issues

Link to [spreadsheet]

Issue	Topic	NHL	Hockey Reference	Natural Stat Trick	Evolving hockey	NHLnumbers		Counting	Ratios	Difference	Smarts	Averages	Percent of total	Heatmaps	Shot locations	Histograms	Correlation	LevelOfDetail	Type
1	Shot attempts -vs- Goals	x	x					x	x	x								Team	Tableau
2	Shot attempts by type	x						x	x									Player	Tableau
3	NHL Standings		x					x	x									Team	Tableau
4	Team luck through shots and goal rates		x						x									Team	Tableau
5	Smart play using the boards; example										x								Comic
6	Smart play passing into space; example										x								Comic
7	Power Play success through duration	x						x	x			x						Team	Visual
8	National Pride	x						x	x				x					Player	Visual
9	Shot attempts & Scoring chances			x	x			x	x					x				Team	Visual
10	Passing lane precision; example	x																	Comic
11	The coaches' whiteboard																		Comic
12	The CAP space					x		x	x									Team	Tableau
13	Killing penalties	x		x	x			x	x						x			Team	Tableau
14	Who do you put on the faceoff?	x						x	x									Player	Tableau
15	SEAHAC2019																		Comic
16	Saves per goal against	x						x	x					x				Player	Tableau
17	Slot Shots				x			x	x						x			Player	Tableau
18	Point split ratios Forwards & Defencemen		x					x	x									Team	Visual
19	Young guns		x					x	x							x		Team	Visual
20	The back of the jersey		x					x										Player	Visual
21	Correlation			x													x	Team	Visual
22	The shootout	x						x	x									Player	Visual



Topic: Shot attempts -vs- Goals



Topic: Shot attempts by type

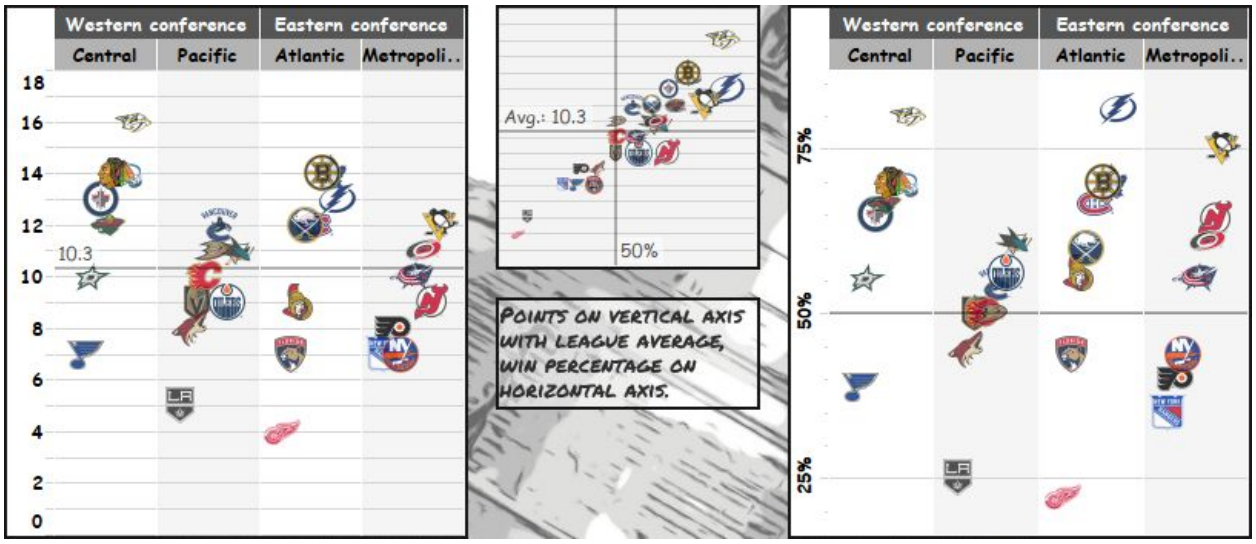
What I tried to look at was not just how many shots a player takes and then compare it to the number of goals, but rather look at all a player's shot attempts, including shot attempts missed, saved and blocked. To this day I do not know how a "broken stick shot" is recorded, but if considered a missed shot that would be included too.

Personally I find it more interesting to see how many goals a players scores considering all his shot attempts, regardless if it is on net. I mean after all, he's still trying to score, isn't he? Just happened to be blocked, a few inches wide (or lots, that's possible too) or hitting the post or bar.

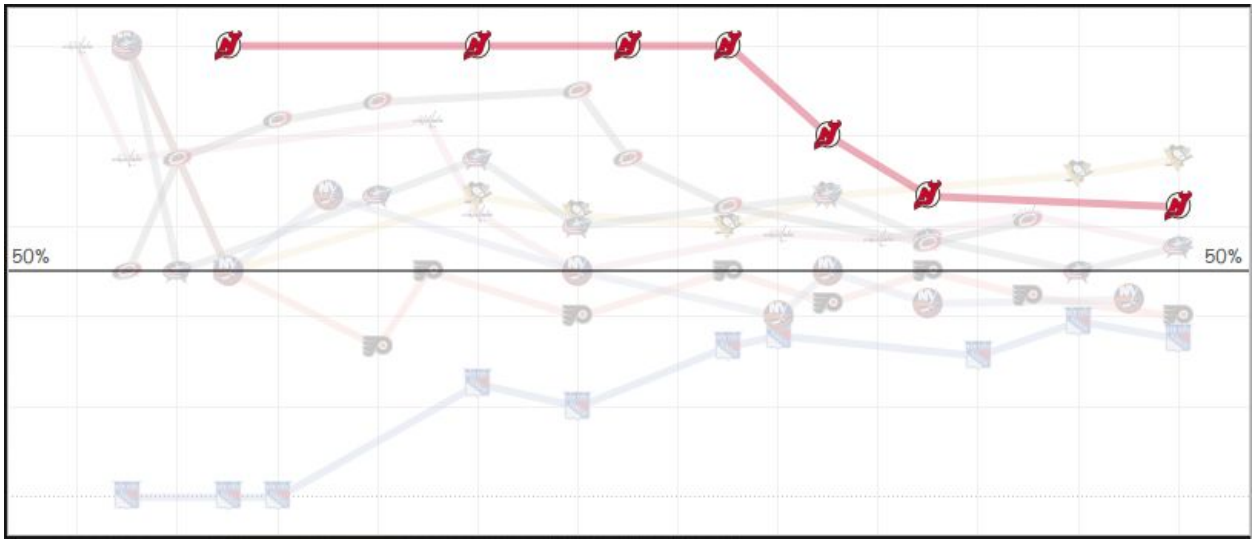
Once the season is going a bit longer I'll add data for the current season once there is a good amount of shots available.

Version 3 of Puck Possessed looks at the standings in the NHL, but not just your typical points, but rather winning percentage. I like this much better as it considers the number of games played, so you get a clear idea of what team is actually doing well.

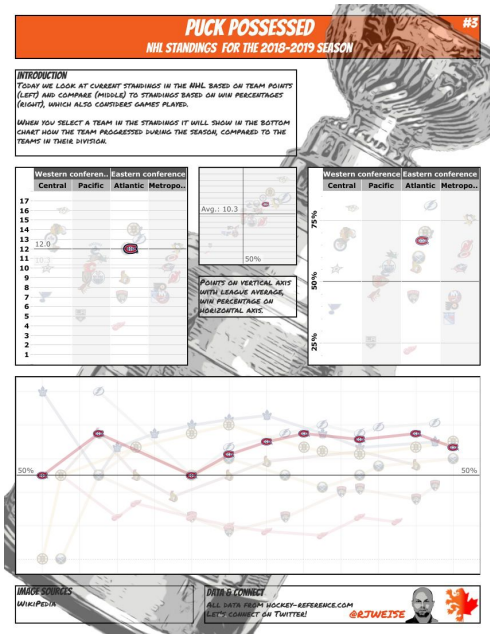
The dashboard first starts with the typical standings table and the winning percentage table, but in between them is a chart that shows both, points on the vertical axis and win percentage on the horizontal axis. As this dashboard updates nightly I have to work with a screenshot here but you can see for example that although the Devils are 4th in their division based on points, they are actually only behind the Penguins when looking at the points they got compared to all points they could have gotten, aka their win percentage.



Playing less games than most of their competition, the Devils started quite well, but with their first game in Sweden they are just a bit behind in games played.



You can find a link to report on **Tableau Public**, where you can select your own favorite team, and explore other team. Have fun!

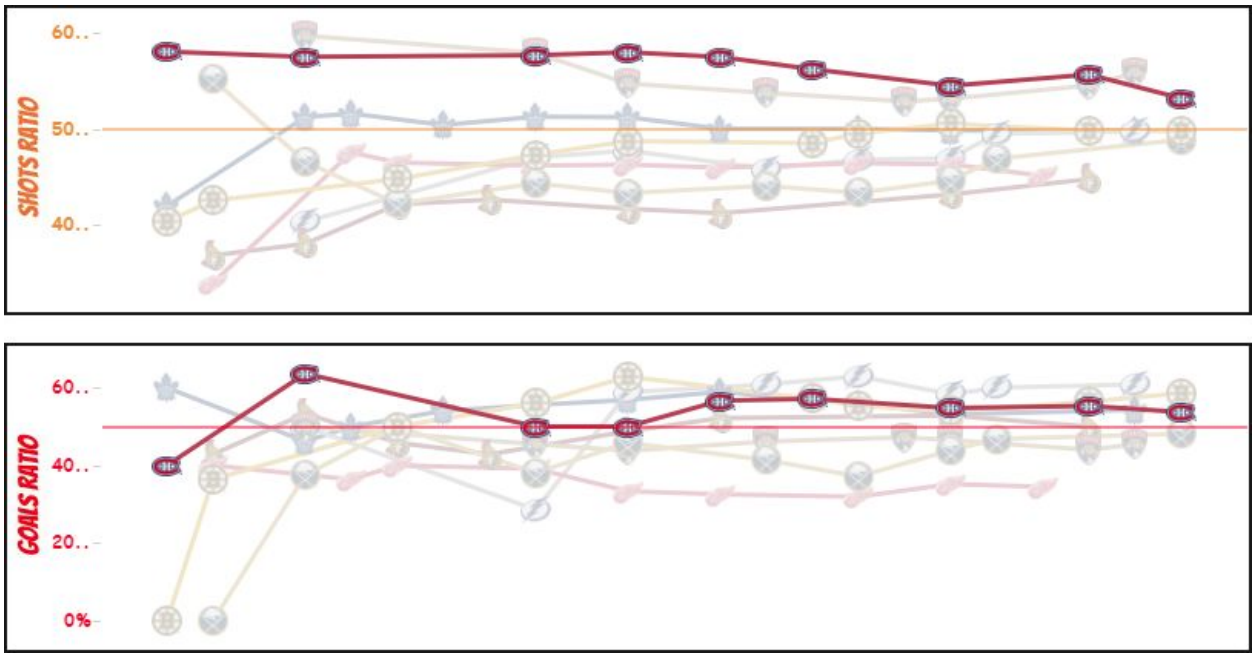


Puck Possessed # 4

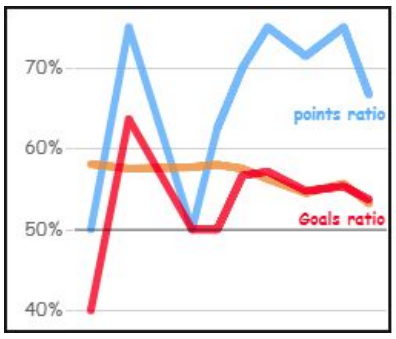
Edition 4 already, and we are looking at NHL team luck, expressed through shot and goal rates. Take Montreal for

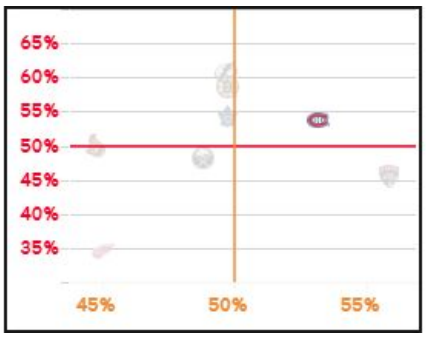
example, a team that not many people are expecting much from this season, other than the "Presse de Montreal". But so far they have doing well, and at the time of writing they have a win percentage of over 60%, closely behind Boston and Toronto (to see current standings based on points and win percentage, check **Puck Possessed 3**).

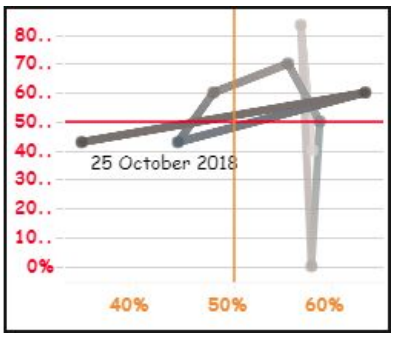
Where other teams show luck, or lack thereof, with a shot ratio and goal ratio that strongly differ, Montreal have actually been both out-shooting and out-scoring their opponents more than 50% of the time.



Unfortunately, both shot and goals ratio have been slightly declining, which has also had it's impact on their win ratio. The last graph on the dashboard, bottom right, follows the Habs on a game by game basis looking at both shot and goal ratio, so see any trends and variation. But other than their game on October 11, where they outshout LA but still were shut-out and lost, and their most recent game against the Sabres, where they lost 4-3 but were out-shot 42-22, they we're in the area you want to be (around 50/50 and to the right and above).

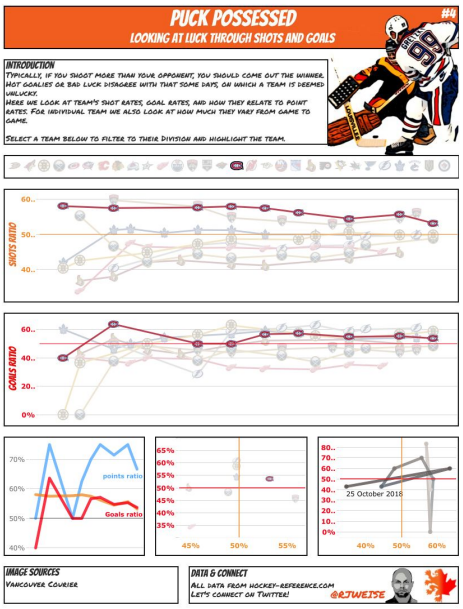






I hope these charts provide some insights, or a different look at the NHL teams. Next up will be some video analysis.

Here's the **link**.



Puck Possessed # 5

Brock Boeser scored a great shot from the circle against Colorado on November 2nd, but I loved the way he got the puck from **E. Pettersson** even more. Not only because it was a great

play, but because it way such a **smart play**, especially considering how young Pettersson is.

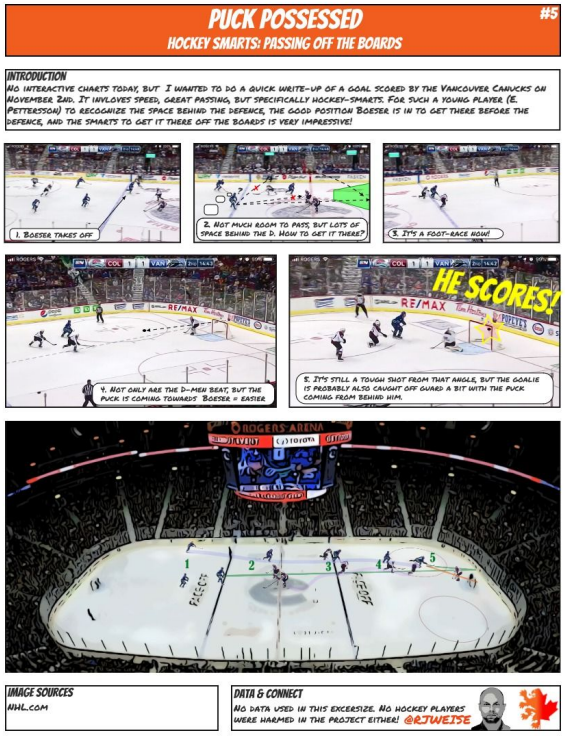
Hockey is a lot about creating good quality shot opportunities, which often means finding open space in close proximity to the net, getting the puck to a team mate in that area, ideally free from any opponents.

As this issue of **Puck Possession** illustrates, Pettersson choose an interesting way to put the puck on Boeser's stick in such a position. I addition, I can only imagine how hard it is for a goalie to see a puck come from more or less behind you, with an unpredictable bounce, especially into the stride of the on-storming opponent.

And for Boeser, despite his skills to receive a puck from anywhere, getting the puck from in front of you while in full stride must making taking a good shot slightly easier as well.

Lastly there is one more benefit I can think of: if the pass does not work out (weird bounce etc.) Boeser would still be fastest to the puck, avoiding icing, and put him in possession in a spot from where he can give a good pass. The other options would like end up in a turn over.

All in all, Pettersson chose the best solution, and in this case it worked out perfectly (unless you root for Colorado!).



Puck Possessed # 6

Puck Possessed # 7

Power play success expressed through the time it takes to score a PP goal

Power play success is typically shown as a percentage: number of power play goals as a percentage of the number of power play opportunities. Nothing wrong with that. But I was curious to see how long a team typically takes to score during the 120 seconds of power play. Yeah, big caveat; see below in the Data section. And pretty soon I realized it would not be as easy as I initially thought. You'd think I would have figured that out by now…

THE DATA

I wanted to use data from the NHL site, since I was (and still am) not aware of any data source available with the kind of information I needed. So I did a couple of things (code available on **my github page**):

- created R-scripts to scrape NHL's play-by-play JSON files for every play and it's time of occurrence in the game


```
{
  "copyright" : "NHL and the NHL Shield are registered trademarks of the National Hockey League.
© NHL 2018. All Rights Reserved.",
  "gamePk" : 2017020819,
  "link" : "/api/v1/game/2017020819/feed/live",
  "metaData" : {
    "wait" : 10,
    "timeStamp" : "20171008_124714"
  },
  "gameData" : {
    "game" : {
      "pk" : 2017020819,
      "season" : "20172018",
      "type" : "R"
    },
    "datetime" : {
      "dateTime" : "2017-10-07T18:00:00Z",
      "endDateTime" : "2017-10-07T20:37:29Z"
    },
    "status" : {
      "abstractGameState" : "Final",
```


;

- created R-scripts to scrape NHL's game summary HTM(L) files to get team strength for goals, who was on the ice, and penalty details

;

- Loaded all this data into a SQLite database (if you are looking for a lightweight SQL environment, hosted freely and accessible from any machine, I highly recommend checking out **https://sqlitebrowser.org/**) and wrote some SQL

scripts to bring things together so to speak (with some help from Google Sheets, believe it or not);
- Loaded the data into Tableau Public, made a number of calculated fields to select the PP goals meeting the criteria mentioned further down, and exported the resulting chart into a stand-alone data source;
- Got the NHL Power play summary data for the selected time frame and combined that with the above mentioned data set

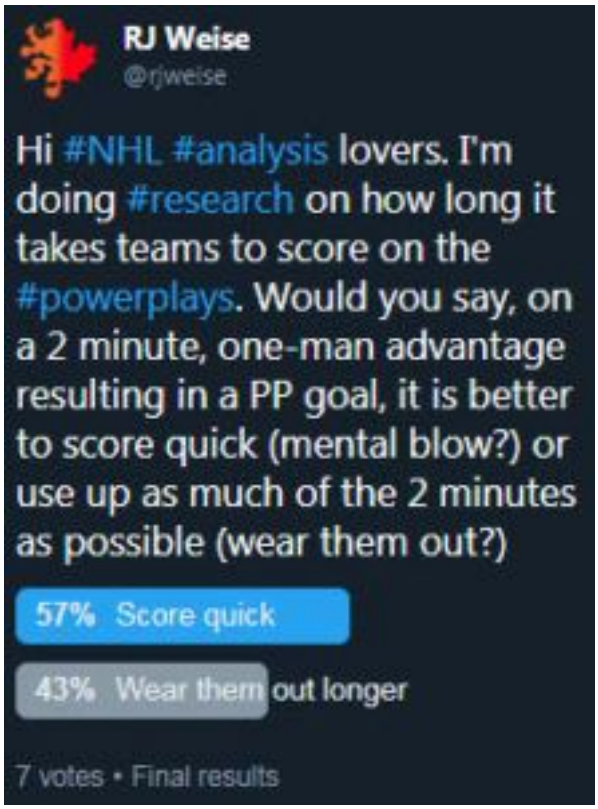
;
- I filtered out only power plays for minor penalties (120 seconds) and with a one-man-advantage to, you know, compare pucks to pucks equally. The time frame is the 2017-2018 season and the first 307 games of the 2018-2019 season (no promises to update this at the end of the season);
- Time to look at the data in Tableau! Quite frankly, looking at the data visually made me think for a minute; I assumed that scoring quickly into the power play is a good thing. But looking at the Penguins here, they took more time and perhaps wore out their opponents more, still scoring on the power play.

RESULTS

So the more I was working with the data, the more I wondered what values would be considered good vs. bad. Not that I think that teams could/would delay scoring on a PP, but is it more beneficial to score at 1:59 of the PP, using up as much time of the 2 minutes as possible and wearing out the opponent, or is scoring 10 seconds into the PP a mental blow to the opposing team, having a bigger impact than the physical wearing out?

I posted a poll on twitter to find out general opinion, and although I appreciate seven folks taking the time to respond,

the sample size is a bit small to come to a conclusive answer. And even still, it would have been pretty much even:

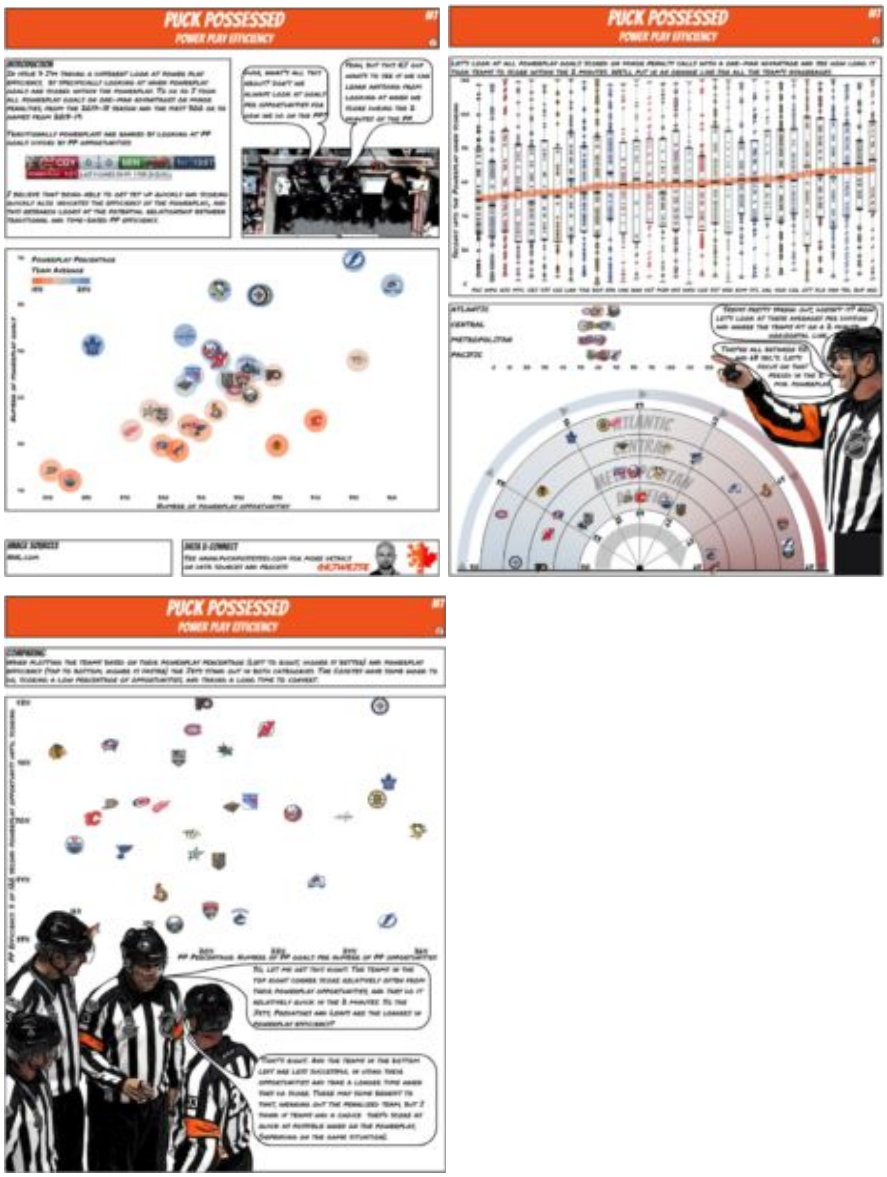


The issue goes beyond one page for the first time, as I am showing traditional PP success with a feature to look at team's PP shot locations, this alternative view of PP success, and how they relate or compare.was not going to fit everything I wanted to show on one page.

Page 1 shows the traditional power play percentage by power play opportunities and power play goals;

Page 2 shows all goals scored in the specified time frame on one-man advantages after minor penalties, called power play efficiency. The main conclusion that these type of goals seem to be randomly and widely spread over the 2 minutes of the power play. The team averages are then displayed, and looked in more detail (all team averages are between 50 and 68 seconds in the power play) to better see the differences;

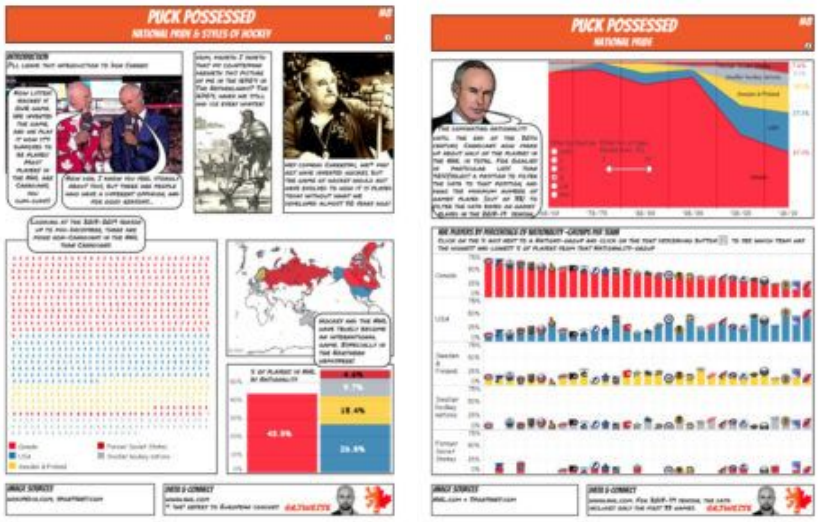
Page 3 shows the combination of power play percentage and efficiency.



The interactive version can be found on [my Tableau Public page](), as usual.

Puck Possessed # 8

National Pride



In Canada we tend to call Hockey “our” game. Although possessing something as intangible as the game of hockey, or any sports game for that matter, Puck Possessed issue 8 looks at National Pride in hockey expressed by players in the NHL representing different “nation groups” in the NHL.

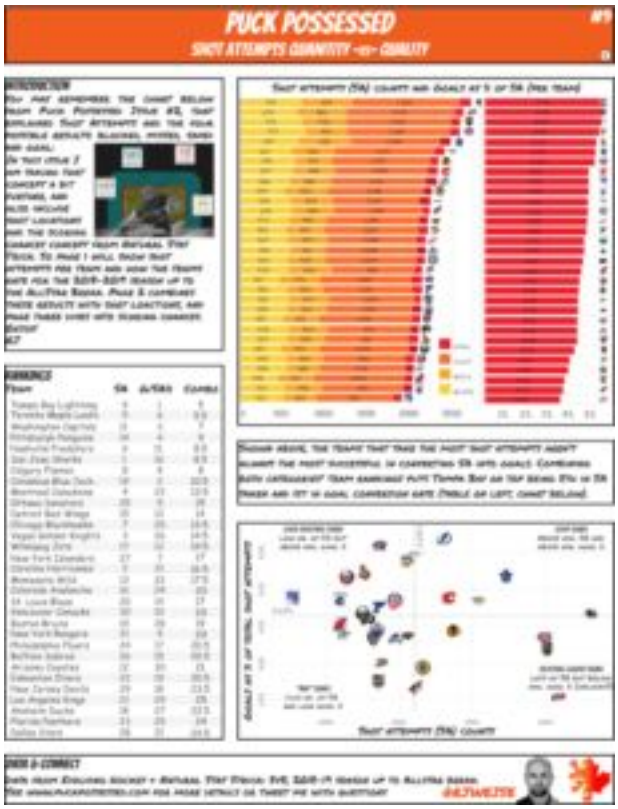
As looking at all individual nations didn’t make much sense, I combined Sweden and Finland, all former Soviet states, and a rest group combining numerous smaller European countries with one-offs like Australia and Bahamas. I should also note that these are based on the Nationalities listed on the NHL web page; I do not know how the NHL deals with dual Nationalities for example, and obviously it also ignores those with a passport from their birth country who actually only lived there their first 3 years of their lives. Initially I also wanted to look at playing styles based on nationality (or continent) but that got kind of diluted by the prior comment, so I left it (for now).

Although both pages are interactive on **Tableau Public**, page two actually has two filters (position and Games Played) that make looking at the chart even more interesting.

Puck Possessed # 9

Puck Possessed Issue 9: shot attempts and scoring chances

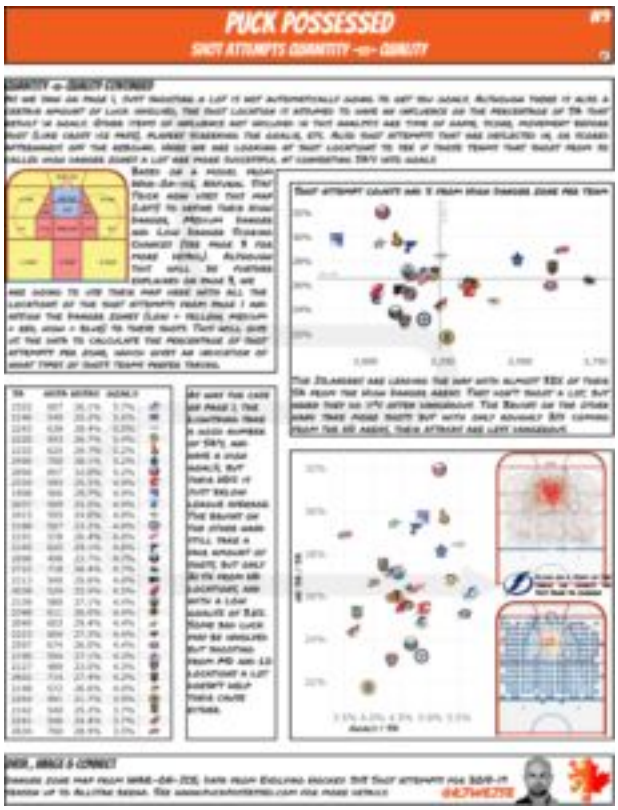
In issue 2 I talked a bit about shot attempts, but the current issue goes deeper, and also looks at, and explains, scoring chances.



The first page of the issue is about shot quality and quality. Is it better to show often or pick and choose only those locations that have a higher statistical chance of scoring? A table compares how teams rank for both shot attempts (abbreviated as SA, not to be confused with shots against) and

Goals as a ratio of SAs. These two rankings are then combined, and not to anyone's surprise, the Lighting come out on top. They still shoot often (ranked 9th out of 31) but also convert SAs into goals most of all teams.

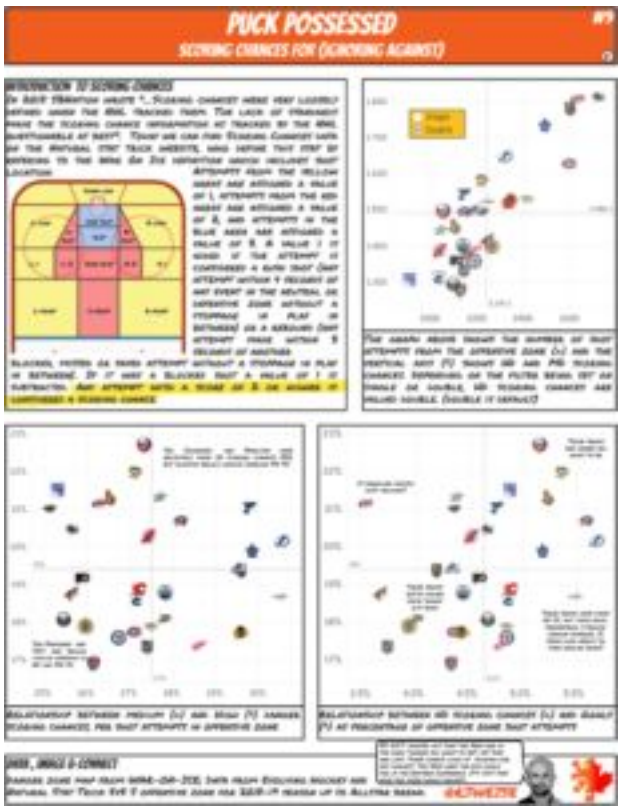
The first bar chart show the same SAs but per event type as tracked by the NHL (blocked, missed and saved shots, and goals). The scatter chart shows the same data again, but also shows how the teams compare with the league averages. It is typically the case that teams that are far from the average will get back closer to the average over time, which could indicate a decline in goal conversion rates for teams like Lightning, Blue Jackets, and Penguins, and the opposite for teams like Coyotes, Kings and Canes. With the Canes shooting well above the league average number of SAs an improvement in Goals% can make them a team to keep an eye on.



The second page goes a bit deeper by looking at where shot attempts are taken from. The assumption is still that just shooting more does not necessarily

lead to more goals. Now we look at the percentage of SAs shot from High Danger areas, further explained on the dashboard. Lastly the relationship between goals per shot attempts, and high danger shot attempts as compared to all shot attempts. Apparently the Islanders take a lot of shot attempts from high danger areas and also do a good job converting them into goals. However, unfortunately for them they produce one of the lowest numbers of shot attempts in the league. To see where they shoot from, click on any team icon and see their shot charts updated.

This is as good a point as anywhere to mention this issue strictly looks at shot attempts for, in other words shots etc. taken by the team of interest. Shot against are ignored here. So teams that do well in these charts and tables don't always do well in the NHL, as they get more shot attempts / goals against than they do themselves.



Page 3 further explains the concept of scoring chances based on work from War On Ice and Natural Stat Trick and compares shot

attempts from the offensive zone to medium and high danger scoring chances. As you will read, this goes well beyond just looking at the location of the shots, and not every shot attempt is a scoring chance, according to this model. Lastly I look at high danger scoring chances compares to medium danger scoring chances and goals.

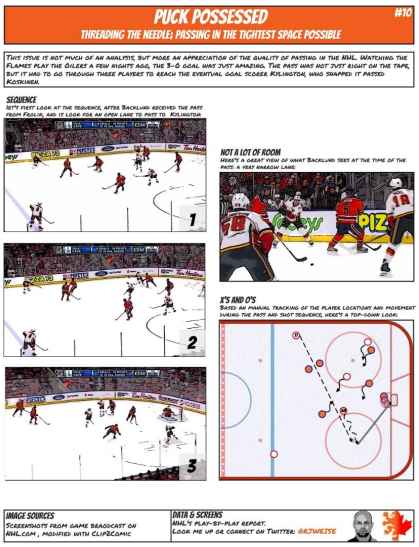
I hope you find this useful and interesting. As usual you can connect with me on **Twitter** or send me any feedback, suggestions or comments through there as well.

the interactive dashboard can be found here: **https://public.tableau.com/profile/rj7974#!/vizhome/PP9ScoringChances_newversion2018/PuckPossessed9p1**

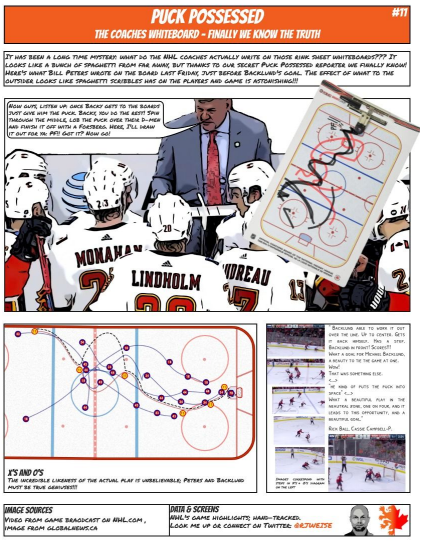
Cheers,

RJ

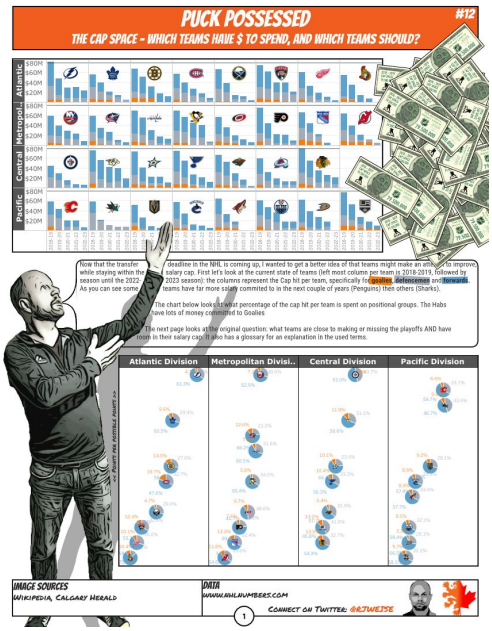
Puck Possessed # 10



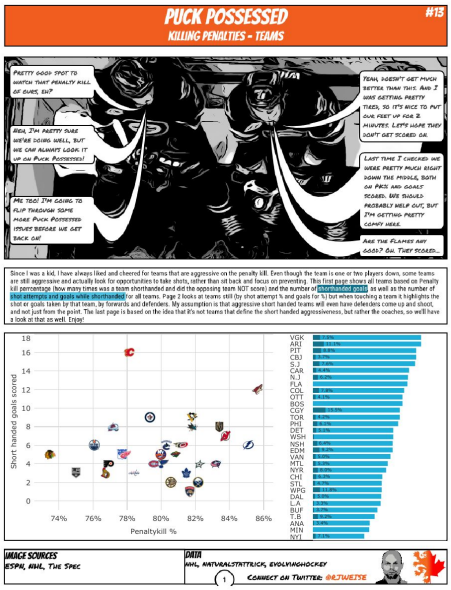
Puck Possessed # 11

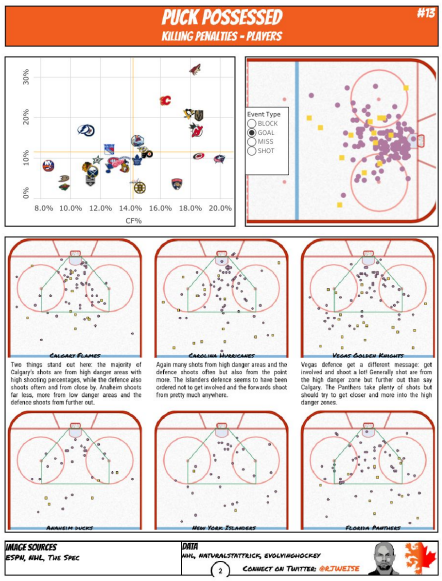


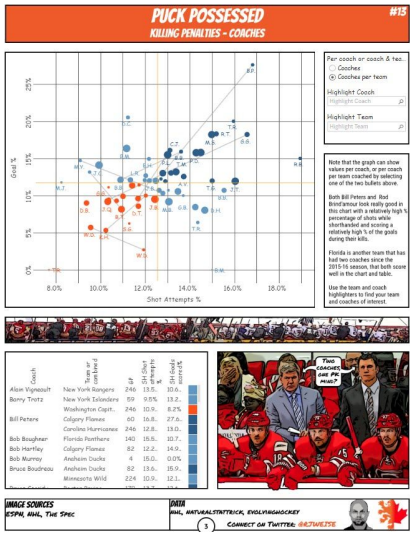
Puck Possessed # 12



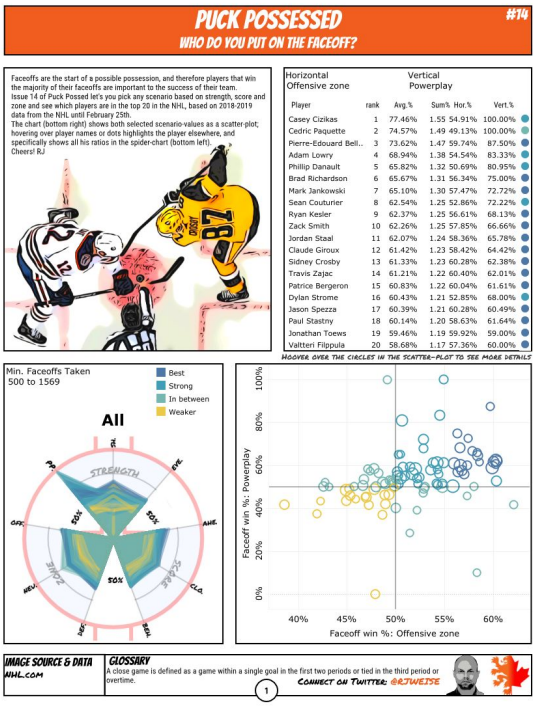
Puck Possessed # 13





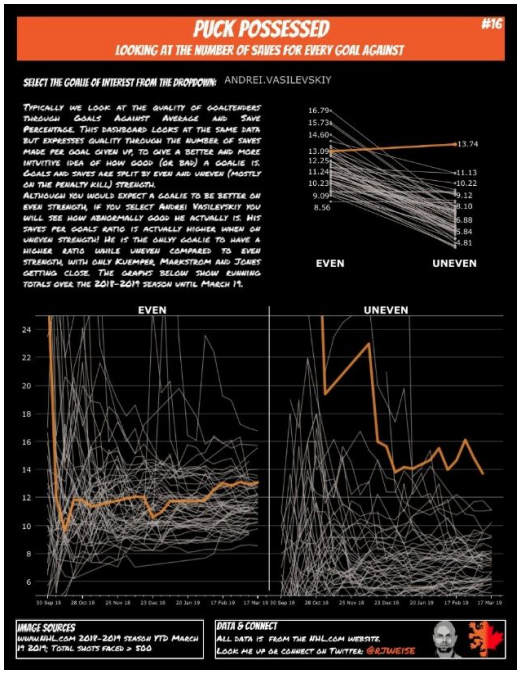


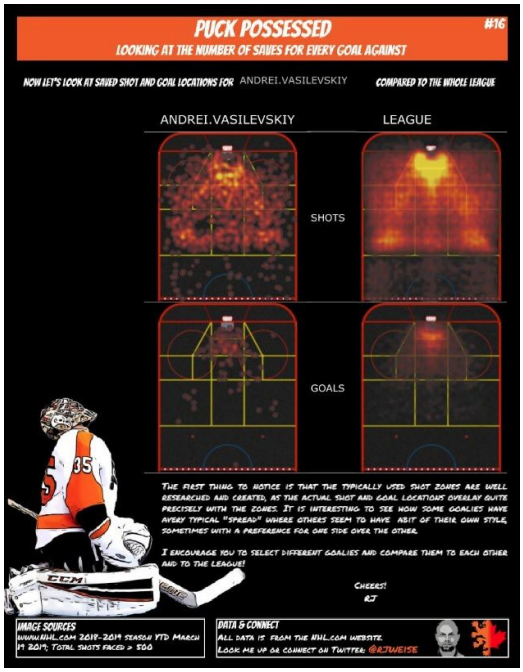
Puck Possessed # 14



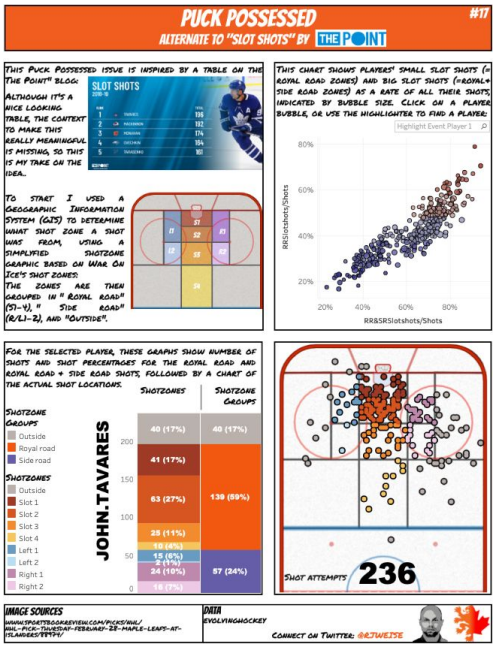
Puck Possessed # 15

Puck Possessed # 16

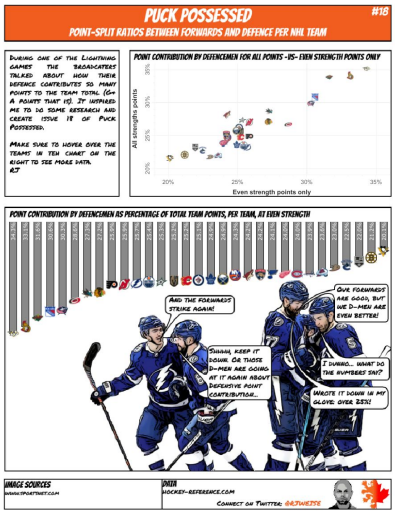




Puck Possessed # 17



Puck Possessed # 18


PUCK POSSESSED
#18
POINT-SPLIT RATIOS BETWEEN FORWARDS AND DEFENCE PER NHL TEAM
DURING ONE OF THE LIGHTNING GAMES THE BROADCASTERS TALKED ABOUT HOW THEIR DEFENCE CONTRIBUTES SO MANY POINTS TO THE TEAM TOTAL (Go A POINT THAT IS). IT INSPIRED ME TO DO SOME RESEARCH AND CREATE ISSUE 18 OF PUCK POSSESSED.
MAKE SURE TO HOVER OVER THE TEAMS IN TEH CHART ON THE RIGHT TO SEE MORE DATA.
AJ
POINT CONTRIBUTION BY DEFENCEMEN FOR ALL POINTS -VS- EVEN STRENGTH POINTS ONLY
All strengths points
Even strength points only
POINT CONTRIBUTION BY DEFENCEMEN AS PERCENTAGE OF TOTAL TEAM POINTS, PER TEAM, AT EVEN STRENGTH
AND THE FORWARDS STRIKE AGAIN!
Shhhh, keep it down. Or those D-men are going at it again about Defensive point contributions...
Ove forwards are good, but we D-men are even better!
I dunno... what do the numbers say?
Wrote it down in my glove over 25%!
IMAGE SOURCES
DATA
HOCKEY-REFERENCE.COM
Connect on Twitter: @AJTWEETS

Glossary

For stat descriptions, see Natural Stat Trick

Sources

ALBERTA CURRICULUM for K-9:
Curriculum: http://www.learnalberta.ca/content/mychildslearning/index.html