ARE THERE CAREERS IN HOCKEY ANALYTICS?

In 1972, Fred Shero of the Philadelphia Flyers hired the NHL's first full-time assistant coach, Mike Nykoluk.[107] Prior to Nykoluk's hiring, every NHL team had just a single coach standing behind the bench wearing a suit, a fedora, and a grim expression.

The league changed slowly at first, but within a generation or two, every team had an entire coaching staff of four to eight coaches, including an assistant coach to help run the bench, a special teams coach to design and run the power play and the penalty kill, an eye-in-the-sky to watch the game from the press box, a goaltending specialist, a conditioning coach, a development coach, a video coach, a special coaching advisor, and more. Plus, nobody wears fedoras anymore.

It's easy to see where I'm going with this. Most NHL teams today have a massive staff of coaches, doctors, trainers, scouts, equipment managers, and so on, but only a single statistician. Often this single statistician will also have other duties and may not have any particular education or experience in the field. Many of them are paid very modestly, play only a minor role in team decisions, and aren't even acknowledged publicly.

Within a generation, expect that to change. Just like with its coaching staff, each team will have an entire analytics department composed of a variety of specialists. As is the case in a few other sports right now, tomorrow's top hockey teams will scout academic institutions that have the most prestigious mathematics and statistics departments, and they'll be armed with big contracts, looking to hire the best and brightest graduates.

The good news is that the Nykoluk moment has already occurred in this field, back in 2014. In what was termed the summer of analytics, about a dozen prominent hobbyists were hired by NHL front offices in the span of a few months.

At the start of the preceding season (2013–14), Eric Tulsky had measured the level of adoption of hockey analytics, and he could find evidence that only 18 of the NHL's 30 teams were using it at all, mostly in very limited terms.[108] For instance, he could only clearly identify four teams that had a single full-time analytics employee, one of whom also had other duties, and three more who used regular analytics consultants.

After the summer of analytics, in which Tulsky himself was hired by the Carolina Hurricanes, Josh Weissbock repeated this exercise.[109] At the start of the 2015–16 season, he could confirm that 26 of the NHL's 30 teams were using hockey analytics and to a significantly greater extent than before. At that point, there were 16 teams with at least one dedicated statistician, most of those teams had several statisticians on staff, and eight of those teams had built an entire statistics department with a director of hockey analytics that reported to the general manager. Weissbock himself was hired as a prospect consulting specialist by the Florida Panthers soon thereafter.[110]

107. Eric Duhatschek, "Hockey's Biggest Shift: Fifty Years of Evolution in NHL Coaching," *Globe and Mail,* March 10, 2017, https://beta.theglobeandmail.com/sports/hockey/from-xs-and-os-to-ipads-the-evolution-of-coaching-in-hockey/article34268774/.

108. Eric T, "Catching Up: The Increasing Pace of Team Adoption of Hockey Analytics," *Outnumbered* (blog), October 21, 2013, http://www.sbnation.com/nhl/2013/10/21/4857618/nhl-stats-advanced-analytics-teams.

109. Terrell Williams, "Hurricanes Hire Eric Tulsky as Hockey Analyst," National Hockey League, August 20, 2015, https://www.nhl.com/hurricanes/news/hurricanes-hire-eric-tulsky-as-hockey-analyst/c-777355; Josh W, "NHL Analytic Teams' State of the Union," *Hockey Graphs* (blog), September 9, 2015, https://hockey-graphs.com/2015/09/09/nhl-analytic-teams-state-of-the-union/.

110. "Florida Panthers Name Richard Pollock Pro Scout and Josh Weissbock Prospect Consulting Specialist," National Hockey League, June 18, 2016, https://www.nhl.com/panthers/news/florida-panthers-name-richard-pollock-pro-scout-and-josh-weissbock-prospect-consulting-specialist/c-886387.

As I write this in the summer of 2017, the adoption of hockey analytics has continued to grow. The NHL launched a new in-house "enhanced stats" platform, all 31 teams employ at least one statistician, leading hockey analytics consulting firm SPORTLOGiQ has contracts with 21 teams, and the Arizona Coyotes even chose statistician John Chayka as their GM.[III]

The growth of hockey analytics is well underway, and many more opportunities will open up in the future. So there *are* careers in hockey analytics, and there will be more. However, who gets these jobs, what do they do, and how do they do it?

Being a long-time pioneer and something of an ambassador in this field, I have been in the unique position to gain a little bit of insight into what's going on behind the scenes. Whether you're interested in a career in hockey analytics yourself or just curious about how these roles are playing out in NHL front offices, I'll do my best to describe what I've seen behind the curtain.

Despite the growing number of opportunities within it, the first step is to think outside the NHL. Everybody's dream is to work for an NHL team, but there are only 31 teams, and many only hire one statistician. Compare that to the average of 500 to 1,000 applications that NHL teams can receive for publicly posted job openings.

Most statisticians get their start outside the NHL. Consider the number of opportunities with the 60 teams in each of the Canadian major juniors and NCAA Division I, plus 132 more teams in the Canadian minor juniors, 36 more in U Sports, and the other 57 North American semi-pro teams between the AHL and the ECHL. Plus, there are also opportunities in women's hockey, the European leagues, and elsewhere.

Furthermore, most NHL teams outsource a lot of their analytics work to third-party agencies, of which there are dozens. In fact, Chayka got his opportunity with the Coyotes by founding his own hockey analytics company in 2010, Stathletes, which now has 50 employees. For many statisticians, careers with one of these companies may prove to be more rewarding than working directly for an NHL organization.

And don't forget that the leagues themselves require statisticians, as do player agencies, and mainstream broadcasters that cover the various leagues and teams. In short, there are a lot of organizations looking for expertise in hockey analytics.

Whether working for an NHL team, in another league, or with a third-party organization, the job opportunities usually boil down to one of three roles: statistical analyst, programmer, or manual tracker. Most opportunities are a combination of the three.

Let's take a closer look at each role, what the job requirements are, what tools and technologies the statisticians use, and how they can help an organization.

Manual Trackers

The first thing any organization needs to do is collect the data, which is why some of the most popular positions right now involve manual tracking. It involves acquiring and classifying NHL data over and above what the official scorekeepers record and gathering even simple data from developmental leagues, which record very little data at all.

These are highly flexible positions that can be either local or remote and either part-time or full-time. They can be with team-level organizations in the NHL or other leagues or with a growing number of third-party companies to whom this work is commonly outsourced. Either way, here's how they work.

▓▓▓ 1. THE TRACKER WILL BE SENT VIDEO OF THE GAMES

Unless the information is being compiled by official NHL scorekeepers or is being used to report directly to the coaching staff at intermission or immediately after the game, the trackers are not required to attend the games and do the work live. Tracking live is

III. Evan Sporer, "Primer: Get to Know the New Enhanced Stats on NHL.com," National Hockey League, February 20, 2015, https://www.nhl.com/news/primer-get-to-know-new-enhanced-stats-on-nhl-com/c-754260; John Matisz, "What We Learned at the Ottawa Hockey Analytics Conference," *Toronto Sun*, May 8, 2017, http://www.torontosun.com//2017/05/08/what-we-learned-at-the-ottawa-hockey-analytics-conference; Jerry Brown, "Coyotes Name Chayka, 26, General Manager," National Hockey League, May 5, 2016, https://www.nhl.com/news/coyotes-name-chayka-general-manager/c-280574260.

exceptionally difficult to do quickly and accurately, unless the events being tracked are few and simple, the tracker is using special technology, or the tracker is particularly experienced.

More commonly, video for multiple games is sent to the tracker, who will complete the assignment on evenings and weekends, usually taking a few hours per game.

2. THE ORGANIZATION WILL SEND THE TRACKER A LIST OF EVENTS TO TRACK

The most common requirement is to track shot attempts, followed by puck touches, passes, or zone entries and exits. However, I have heard of at least 50 different event types over the years, including screens, stretch passes, plays up the end boards, hits that resulted in a turnover, and pretty much anything else you can think of.

Tracking the most common events is usually pretty straightforward and requires very little experience to complete accurately. However, some events require considerable hockey knowledge to properly interpret and record, and they are therefore usually assigned to former players, coaches, and scouts, along with careful instructions about how these events are defined and how to record them.

3. THE ORGANIZATION WILL SEND THE ANALYST SOME SOFTWARE USED FOR VIDEO TAGGING OR PERHAPS A LAPTOP WITH THE REQUIRED SOFTWARE ALREADY INSTALLED

In some cases, the organization wants more than just a count of the events and who was on the ice. They want an actual video file tagged with the time of each event. That allows the coaches to quickly study specific plays and matchups.

Video tagging requires a little bit of training and specialized technology. Leading examples of such software include Steva Sports Software (stevasports.com), XOS Digital (xosdigital.com), Gamebreaker (Sportscode) by Digital Tec Solutions (dtsvideo.com /gamebreaker-plus), PUCKS by Sydex Sports Software (sydexsports .com/html/hockey.html), and dozens of others.[112] Licences are quite

expensive and possibly unavailable to those outside NHL organizations, so the software must be provided by the team. To my knowledge, there is no industry standard for video tagging, and some organizations use their own proprietary software.

Depending on the exact technology, this software allows the trackers to watch a video of the game and to quickly tap the screen to record the type of event, along with additional information, such as the time it occurred, where it occurred on the ice, and the jersey numbers of the players involved. It's quite remarkable to watch how quickly and accurately a seasoned expert can tag a game using sophisticated software like this.

Unless video tagging is required, most manual trackers will do the work by hand either with pen and paper, directly into a spreadsheet, or by using custom software developed for just this purpose.

4. THE TRACKED DATA IS RETURNED TO THE ORGANIZATION

The end result is usually a formatted list of times, events, locations, and jersey numbers that is produced manually or is exported from the software in a variety of formats. If video tagging was required, then the end results will include a tagged version of the video itself.

If this was done for an NHL game, then the results can be integrated into the league's official shift charts and play-by-play files.
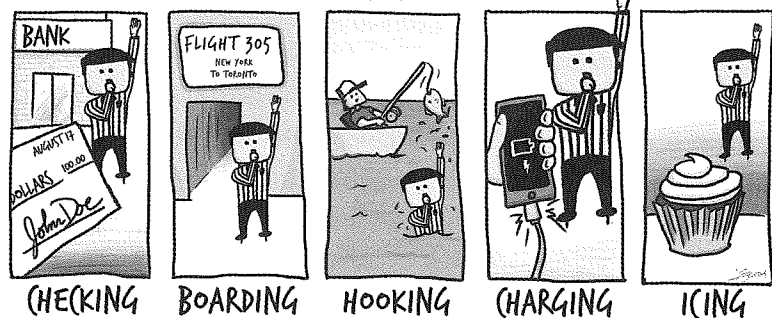
And that's about it.

Since the work is flexible and requires no experience in statistics or programming, it is a great first step into the world of analytics. In time, the manual tracker can be involved in the subsequent analysis of this data and/or the programming tasks required to store it, manipulate it, and create reports.

To get started in a career in manual tracking, the first step is to find (or create) and get comfortable with a convenient process for tracking game data quickly and effectively. This is often built into a spreadsheet, but there are a number of cheap applications out there. Manual tracking can even be done with XOS, and it pays to learn the video tagging technologies anyway.

Next, learn how to track as many different kinds of events as possible. Even the simple events aren't as straightforward to record

---

112. At the time of writing, TVTI, which was the video provider for PUCKS, had been bought by Stats LLC, which is making their own version of PUCKS called STATS. PUCKS now has a new video provider.

REFS IN THE OFFSEASON

CHECKING   BOARDING   HOOKING   CHARGING   ICING

as most people would expect, since hockey has a way of creating so many wild situations.

Try to get some experience by contacting local junior, university or college, and minor-league teams in your area, even offering to work for a low wage or on a volunteer basis. Another way to get your feet wet is to consider joining an online community that does manual tracking for independent research purposes, like Ryan Stimson's Passing Project.[113]

Even if it doesn't net you the big bucks, this experience will help prepare you for a meeting or an interview with solid examples of your work. I also recommend measuring the speed and accuracy with which you record various types of events and how fast you can tag them using the different types of software.

When looking for job openings, don't just look at the team or league level. In many cases, the work will be outsourced to third-party organizations. One of the more famous cases is Stathletes (stathletes.com), based out of St. Catharine's, Ontario. They were co-founded by John Chayka, who went on to become the youngest NHL GM in history, with the Arizona Coyotes.

Hockeytech (hockeytech.com) is another big player. They were founded by former Panthers CEO Stu Siegel in 2013, based on corporate entities that go back to 1998. They have a variety of services and clients in virtually every North American league, including every NHL team.

And then there's PowerScout Hockey, founded in Ottawa by Marc Appleby. They were one of the earlier manual tracking companies known to work in the NHL and were a pretty big player for years. However, I haven't heard much about them since 2014, and their website is long gone.

There are also a number of small manual tracking companies launched by former NHL players. Two good examples include Brad Werenka's Calgary-based Truperformance (truperformancehockey.com) and Drake Berehowsky's StatsTrack (statstrack.ca) in Toronto. If you live in or near an NHL city, check around to see if there are opportunities to learn and contribute.

If not, many prominent hobbyists have chosen to co-found their own manual tracking company, including Jennifer Lute Costella and LCG Analytics (lcganalytics.com) and Garret Hohl, whose Vancouver-based company HockeyData (hockeydata.com), which focuses on the AHL, landed a contract with the Washington Capitals in May, 2017.[114]

And, the list goes on. There's also Bench Metrics (benchmetricshockey.com), Dark Horse Analytics (darkhorseanalytics.com), and InStat (instatsport.com/en). There could be as many as 100 of these companies. Almost without exception, they also offer analysis and reporting software and services, so they should therefore be of interest to statisticians and programmers, too.

Lastly, manual tracking isn't the only way that teams are getting their data. In the future, video technology will play a larger role. Due to the cost of the special equipment and software, video is far more expensive than manual tracking. However, once it is automated, additional games and countless events will be tracked at virtually no additional cost, and accurate results will be ready within seconds. If enough data is being tracked, or the expense is being shared by enough organizations, then video technology will eventually be as affordable as manual tracking.

---

113. You can support Stimson's Passing Project at his Go Fund Me page, www.gofundme.com/between-the-lines.

114. Randy Shore, "Vancouver Entrepreneurs Net Analytics Contract with NHL's Capitals," *Vancouver Sun*, May 10, 2017, http://vancouversun.com/news/local-news/vancouver-entrepreneurs-net-analytics-contract-with-nhls-capitals.

There are a number of companies who have been introducing this technology to hockey, including PowerScout Hockey, SPORTLOGiQ (sportlogiq.com), Iceberg Hockey (http://iceberg .hockey), Keemotion (keemotion.com), Pixellot (pixellot.tv), StriveTV (strivenetwork.tv), and Hawkeye (hawkeyeinnovations. com), which is also used for coaches' challenges in the NHL.

As the technology evolves, video tracking can also improve the accuracy of the data. That trend has already been observed in other sports, like with PITCHf/x and FIELDf/x in baseball and SportVU in the NBA. Even when multiple special cameras are installed, hockey presents a few interesting challenges for video recognition, like when players get tangled up in a puck battle or when the puck gets obscured by the boards, a player, an official, or the net. That's where augmenting video camera technology with chips in the pucks and/or equipment can come in handy, albeit at even greater expense.

As observed at the 2015 NHL All-Star Game and the 2016 World Cup, using Sportvision technology in the pucks and player tags can provide precise location information several times per second and can even track the angle and speed of the players and puck. Interestingly, hockey was the very first sport to use this technology, which was initially known as FoxTrax (with the infamous glowing puck), all the way back in 1996. Acquired by SMT in 2016, hockey is now listed under the "other sports" category on the company website (sportvision.com).

Despite the growing success of these and other technologies, they are unlikely to ever drive manual trackers out of business. There will always be many leagues for which it just isn't cost-effective to pursue those options, and there will always be specialized events that temporarily fall outside the bounds of what a camera or computer can offer.

Data Scientists

The role of the data scientist is the easiest to describe, because it's very close to what we do in books like these. We start with a question, form a reasonable hypothesis, carefully gather the data, craft and/or identify appropriate measurements, devise tests for our hypothesis, calculate the results, interpret them, and then express the results in a way that makes sense to everybody. In a nutshell, that's exactly what a data scientist does for a hockey team.

Surprisingly, only a few teams invest in data scientists. The others are content to outsource this work or to hire an entry-level analyst who creates reports that merely apply the analysis done in the public sphere.

In time, more organizations who are looking for an edge will hire and retain statisticians and other innovative thinkers of their own. Over the years, we have actually been sneaking some of the more sophisticated terms and techniques used by statisticians into these books, especially in the opening chapter of *Stat Shot*. Long-time readers might actually understand this world better than they realize.

For example, let's consider a case where the statistician's input is being requested for a specific draft pick, a trade, or a contractual decision. The statistician might be a full-time employee, part of a third-party organization, or perhaps just a consultant being engaged on this one matter.

Let's imagine that it's the summer of 2010, you're a statistician for the Montreal Canadiens, and the team's new GM Pierre Gauthier asks you if he should keep Carey Price, who had a career save percentage of 0.912 at that point of his career, over Jaroslav Halak, who had 0.919.

A hobbyist would buy the latest version of *Hockey Abstract*, download the latest data from a few websites, apply what they have learned, and create a good report for Gauthier. There's nothing wrong with that (especially the part about buying the latest version of *Hockey Abstract*), since that contribution would certainly be of good value to the organization. However, statisticians can take it a step further by innovating custom metrics and producing results of greater statistical certainty.

It all starts with tests. Statisticians run a lot of tests, which is exactly how this hypothetical situation would first be approached. Here, the statistician could test the hypothesis that Halak's superior

skill resulted in the higher save percentage versus the so-called null hypothesis that it did not.

Of course, statisticians never deal in absolutes. Statistical analysis is all about probabilities, not certainties. The statistician would have to select a level of uncertainty that is acceptable, which is called the alpha level. Typically, this is 0.05 or 5%.

Establishing that level of uncertainty means calculating the p-value, which is the probability that the null hypothesis is true. In other words, it is the probability that the observed difference in save percentage could have occurred without Halak possessing greater skill. Generating that p-value usually means conducting a sample test of the data, determining the degrees of freedom, calculating the standard error rate, and building a distribution curve. Explaining these terms is outside the scope of this book but don't worry, there won't be a test on any of this.

If that p-value is less than the chosen alpha, then the statistician can reject the null hypothesis and report to Gauthier with 95% confidence that Halak is better than Price. If this later turns out to be one of those 1-in-20 cases where the numbers were wrong, then that is termed a Type I error. When getting into an argument with a statistician on Twitter, I highly recommend you throw that term at them, since it really gets their goat.

In hockey analytics, Type II errors are far more common. Those occur when the p-value exceeds alpha, which means the null hypothesis can't be rejected, even when it is false. In our example, it means that the data didn't definitely indicate that Halak was better when he actually was. While that obviously didn't occur in this particular case, it often happens when there isn't enough data or the data just isn't good enough.

For example, in *Stat Shot* we found that "Braden Holtby, in second place, has a lower bound that exceeds the upper bounds of only two goalies, journeyman backup Curtis McElhinney and Jonas Gustavsson."[115]

That means that we couldn't reject the null hypothesis when comparing Holtby's career even-strength save percentage to almost any active NHL goalie. Arguing that factors other than skill can produce the difference between Holtby's 0.929 save percentage and Ondrej Pavelec's 0.917 is a great example of a Type II error.

If it sounds complicated, well, it is. Plus, statisticians wouldn't even be using something as simplistic as save percentage in a situation like this. Instead, they would be using their own data sets, and multivariate metrics of their own design to measure the meaningful and persistent qualities of a goalie's skill, and then they would use them to compare Price and Halak not just against each other, but against all goalies. In fact, that's exactly where we landed at the end of that goaltending chapter in *Stat Shot*, with a rather comprehensive and sophisticated model of all goalies.

Furthermore, that statistician certainly wouldn't go to Gauthier with talk of p-values and Type II errors, just as we generally don't include those terms in books like these. These books are mostly written for the enjoyment of mainstream fans and hobbyists, and the professional statisticians among you have simply learned to tolerate my oversimplifications (including those previously mentioned). Reports for management have to be filtered and crafted similarly.

At first, most data scientists do all of their work in Microsoft Excel because it is popular, is easy to use, and requires no experience in either programming or statistics to learn and use, unless it is supplemented by an extension like Visual Basic or SigmaXL. According to a survey conducted by Ryan Stimson, 73% of the field use Excel as their primary tool.[116] Even if most statisticians eventually move on to more sophisticated technologies, Excel will always remain a useful technology with which to share one's findings with the rest of the community and/or the front office.

Second place in Stimson's survey went to R, which is gradually becoming the standard in this field. I highly recommend it for hobbyists who want to take their work to the next level, whether or not they are seeking a career in hockey analytics.

What is R? Besides being a very nifty letter with which to spell first names, R is a programming language. Rather than clicking on

115. Rob Vollman, "Who Is the Best Puck Stopper?", *Stat Shot*, 2016, 178–179.

116. Ryan Stimson, "Hockey Analytics Survey Results," *Hockey Graphs* (blog), February 9, 2017, https://hockey-graphs.com/2017/02/09/hockey-analytics-survey-results/.

Excel cells, statisticians do their work by typing commands into the R console and grouping sets of commands into programs. It has a bit of a learning curve for those without any programming expertise, but pretty soon statisticians are able to complete the same tasks they were doing in Excel in a fraction of the time and with minimal effort.

If you're interested in doing a little bit more with statistics, then get started by downloading and installing R, which is free. [117] It comes with the RGui console, but you may want to download and install RStudio, which has a graphical user interface (GUI) that's easier to use and has a few more features. [118]

As for data, R can actually open all of your Excel files as well as any files in the formats used by other statistical tools and technologies. If you're just starting out, then you can grab the data from my website (hockeyabstract.com/testimonials). Load the data using the library command, and then use the read command to open it. Next, start by using the head and summary commands to make sure everything's working and to get your first look at the data set in question.

To start using and analyzing the data, there's no need to be an experienced programmer, since one of the great advantages of R is having access to a lot of great libraries and packages that can manipulate the data, do all the number crunching, and create visualizations for you. For starters, download Michael Lopez's list of the 10 most useful R packages for sports statisticians, to which I would also add NHLScrapR to parse NHL game files. [119] Also, if you're going to be in the public space, then it can be helpful to create a Github account of your own (github.com), in order to organize and share all of their code and solicit assistance and advice from others.

While R is gradually becoming the standard, there are many alternatives. For instance, APL (or J) and MATLAB are very similar statistical programming languages. However, the former is quite

---

[117]. The R programming language can be downloaded at www.r-project.org/.

[118]. RStudio can be downloaded at www.rstudio.com/products/rstudio/download/.

[119]. Michael Lopez (@StatsbyLopez), "Here's a list of #rstats packages, catered to those doing statistics in sports," Twitter, November 8, 2016, 7:38 a.m., https://twitter.com/StatsbyLopez/status/796014026094481408; the NHLScrapR package can be downloaded from the War on Ice Github page, https://github.com/war-on-ice/nhlscrapr.

antiquated, and the latter can be expensive and is more common in academia or other fields. They will absolutely do the job and are a huge step up from Excel, but using these tools makes it a bit more difficult to collaborate with others.

Since it's impossible to know which tool or technology a team is using—or their third-party partners or various independent consultants—collaborating with others might mean having all the options installed on your laptop and acquiring a basic working knowledge of each one. In addition to R, APL, and MATLAB, that might include:

- Minitab, which is extremely simple and therefore popular in academia for instructional purposes. Nobody really uses it for their day-to-day work, but everybody knows it.
- SPSS, which is IBM's highly sophisticated analytics software whose popularity is stunted only by its expense. For those who can afford it, it's the most popular alternative to R, in my experience.
- SAS, or Statistical Analysis System, which is a collection of software tools used for a wide variety of applications in the field of math, data, and statistics.
- Stata, which is very established statistical software that is far more popular in other fields than in sports, much like MATLAB.
- Gretl is a free open-source statistical package, with a GUI that can actually be used with other languages, like R and Stata.
- SAP technology, which is an enterprise resource planning software with an expensive and complex analytics solution. It is used by the NHL and virtually nobody else in this field.

Data scientists may also have a variety of more general-purpose programming languages on their laptop, like Python and SQL, which we'll discuss in the next section.

Given that analytics is used extensively in other fields, such as sales and marketing or business development, you may already have access to some standard tools and technologies, including the licences to use them. To be clear, most organizations don't have any

overlap between the data analysts working in their different departments, but there's no reason that they can't share the technical infrastructure and IT staff.

Regardless of which technology you use, accuracy is essential, especially in data collection. Even if the data is acquired from third parties or manually collected by other people in the organization, the data scientist has the responsibility to organize these massive data sets, to verify their accuracy, to identify and remove any bias, and to become thoroughly familiar with every scrap of information, including their relationship with the scoring and preventing of goals and their year-to-year persistence.

Sometimes data scientists are criticized because, in some people's minds, only wins and losses matter, so we should just toss all the fancy numbers away. However, wins and losses being the only thing that matters is actually the foundational argument of the entire field. In fact, it is the first law of hockey analytics.[120] If a fancy number can't be shown to have a reproducible relationship with wins and losses, then it is tossed aside. That's the entire point of the field, making that criticism one of the field's more bizarre ones.

Misunderstandings like these create the need to communicate and express the results in a meaningful and actionable fashion, whether that's in reports, emails, or presentations. This is especially true in these early days, when there is not likely to be as good a foundational understanding of hockey analytics throughout the organization nor a common language. In many organizations, there's also a degree of skepticism that can go well beyond what is healthy.

As for experience playing hockey, that isn't strictly required because most organizations already have that in spades. However, statistical data can be highly prone to misinterpretation, and it pays to have a solid grounding in how the sport is played.

As for pay, statisticians average about $85,000 per year in the United States and about 50% more in specific sectors or particular geographic locations. In hockey, they may make only half as much. Nobody goes into this field for money or fame; there isn't very much of either one.

120. Alan Ryder, "The Ten Laws of Hockey Analytics," Hockey Analytics, January 2008, http://hockeyanalytics.com/2008/01/the-ten-laws-of-hockey-analytics/.

Programmers

Surprisingly, those who have been most successful at finding work in hockey analytics right now are those with programming skills.

Dating back to Darryl Metcalf, who built the Extra Skater website before being hired by the Toronto Maple Leafs on August 19, 2014, those who can build databases of the field's leading research are perceived to have equal or greater value than those who created such stats in the first place.[121] Even a long-time brilliant hockey statistician like Andrew Thomas wasn't hired until he built the War on Ice website to share his and the rest of the field's latest developments with the general public.

Surprisingly, the programming positions are the ones that I know best, since that's what I used to do for a living. Prior to doing hockey analytics full-time, I spent 15 years in a variety of fields doing programming, technical support, and project management for large data-driven applications. This background has also given me particularly keen insights into what these programming positions entail and what teams are looking for.

The primary responsibilities for a programmer in a hockey analytics department boil down to the acquisition, storage, manipulation, and retrieval of data. In some cases, these are full-time, on-site positions, and in other cases they are done part-time and remotely, or they are outsourced to another company entirely.

Every hockey organization will eventually build a database of hockey stats. Virtually all organizations will already have databases for their other business departments, like human resources, accounting, marketing, and so on. Often, the hockey analytics department will share the same infrastructure, underlying IT staff, and software licences as the rest of the organization.

Building that database and populating it with information will be one of the programmer's key responsibilities. The biggest challenge is that the data will come in a variety of different formats,

121. James Mirtle, "Leafs Go Big on Analytics with Creation of Stats Department," *Globe and Mail*, August 19, 2014, https://beta.theglobeandmail.com/sports/hockey/leafs-go-big-on-analytics-with-creation-of-stats-department/article20129261.

depending on whether it comes from the league, is scraped from the websites mentioned throughout this book, is acquired from third-party data providers, or is manually gathered by the team itself. All of this information must be acquired in an automated fashion and stored in the database in a common format.

Even once the data is converted to a common format, there's often a lot of work that needs to be done to clean it up. That means looking for missing or incorrect data or pieces of data that conflict with others. Just to take one simple example, there might be one data source that lists a player as Matt Benning while another lists him as Matthew Benning. Sometimes that change will take place in the middle of the season, and there will be two data sets for a single player that need to be merged.

Programmers also want to avoid having redundant information in the database. What if they don't add up? For example, goals and shots will be stored in the database, but shooting percentage will not because it is simply a different representation of the other two statistics. Having all three pieces of information in the database could lead to some sticky situations, trying to figure which data is accurate whenever shooting percentage doesn't exactly match goals divided by shots.

That's one situation where having at least a basic understanding of the field is handy. Shooting percentage may be a simple enough statistic for everybody to wrap their walnuts around, but what about GVT or expected goals? Even if the team has a few data scientists to help out, the programmer will still need to have a working knowledge of stats and hockey in order to program the various calculations properly.

Finally, there's the task of making all of this wonderful information available to the rest of the organization. Programmers can build websites like Extra Skater and War on Ice, and they can also produce automated reports that are sent directly to the laptops, browsers, phones, or tablets of coaches, scouts, and GMs. New features are requested all the time, but the programmer is often given a lot of freedom as to how it's done.

To complete these tasks, programmers use a number of tools. To highlight just the four most common ones:

- The data is normally acquired using technology like Python, which is a scripting language ideal for scraping data off the internet and manipulating it for further analysis.
- There are alternatives, like C#, which is the spiritual successor of the technologies I used to use. It is a generic but very popular and powerful programming language that can build almost anything from scratch.
- To store the data into the database, programmers use SQL, which is the Structured Query Language used in databases like Microsoft SQL Server to store, manipulate, and retrieve data of any kind. I used this technology quite extensively, but for a database called Oracle. I even had a very popular website in my previous career (thinkoracle.blogspot.com, which is still online) and was the technical editor of one of the best-selling Oracle SQL textbooks of the time (*Oracle SQL by Example* by Alice Rischert).
- To present the data, there's Tableau, which is arguably the best data visualization software out there for making charts and graphs.

As for compensation, if you thought the pay cut that data scientists have to take in order to work in hockey is extreme, then you really don't want to know about the programmers. The good news is that the flexibility inherent in these positions often means that programmers can work from home and on their own schedule.

Career Advice

Regardless of which position you may want to pursue, my best advice is to learn the fundamentals, don't be afraid, find and trust your voice, and stay focused on your passions.

In the past, my own sense of modesty caused me to resist requests for advice—who the heck am I? But as the field grew and everybody's successes began to accumulate, I started putting a little bit more thought into which of our experiences might be of value to those who are just getting into hockey analytics as a possible career.

At first, my approach was to identify and repeat the best advice that I had received, rather than come up with anything of my own. As I wrote in *Hockey Prospectus* just prior to the 2014–15 season, the best advice I ever received came from a rather unexpected source: a stand-up comedian named Darryl Lenox. His amateur comedy workshops taught local comics how to find and trust their voice, how to keep coming up with good ideas, and how to build their audience.[122]

Even though Lenox's advice was geared toward the stage, it translated well to my glorified hockey stats hobby, too. Rather than try to emulate other stand-up comedians that we admired, his most important piece of advice was to find our own voice and stick to it. The sooner you do that, he argued, the less time you waste developing the wrong talents for someone else's audience.

In our world, that means that we can be inspired by hockey statisticians whose work we admire, but we need to find our own way of approaching this game. We need to find the specific areas that interest us and what approach we'd like to take. Here are 12 questions that can help you figure that out:

- Is this meant to be a casual hobby, or do you want to work in a front office or in mainstream media or elsewhere?
- Do you want to stick to writing, or do you want to make appearances on radio, TV shows, and podcasts?
- If you're a writer, are you interested in blogs, formal online articles, magazines, or books?
- Is your intended audience coaches and front offices, the mainstream fan, the fantasy pool enthusiast, the hard-core statistician, or someone else?

- Is your natural style to muck it up and argue against bad ideas, or to shine a light on the great ones?
- To what extent do you prefer tables of figures, visualizations, or the written word?
- Do you see yourself developing new statistics and innovations, or finding applications and/or new presentations for what's already out there?
- Can you write code to parse data from websites and game files and create automated reports?
- Do you want to create and host your own statistical database website?
- Do you have any interest in recording your own statistics, like zone entries or passing data?
- Is there a specific team you want to cover, or do you want to cover the NHL as a whole or perhaps specialize in another league altogether?
- Do you want to cover the entire game, or do you want to focus on an individual area, such as goaltending, faceoffs, injuries, or special teams?

Even once these questions are answered, it can still take time to find and refine that voice. In my case, I spent an invaluable 2013–14 season writing for Bleacher Report and investing time reading all of the comments to figure out how best to phrase my ideas for the average fan, including those who were new to and skeptical of statistical hockey analysis.

All that interaction with my readers also helped build my readership. Lenox described comedy as a competitive and individual business, not a collaborative one. Sure there are lots of people who will support you and help refine and promote your work, but ultimately a comedian will always find himself alone on a stage with nothing but a microphone, trying to be funnier than the next guy in order to land his next gig. Likewise, in hockey analytics, it will be your name at the top of an article, with only the quality of your own analysis helping you score the next opportunity.

To that end, Lenox's advice was to create your own market.

122. Rob Vollman, "So You Want to Be a Hockey Analytics Writer?," *Hockey Prospectus* (blog), September 5, 2014, http://www.hockeyprospectus.com/so-you-want-to-be-a-hockey-analytics-writer/.

When he comes to town, the comedy club already has an audience and books him only because of the additional people he will bring in. If he doesn't bring his own audience, then they will book someone else who can.

Don't rely on a website, radio station, front office, or publication to promote your work or to provide the audience for you. Cultivate it yourself. Reply to emails, respond to tweets, speak at conferences, go to games, and get engaged in the comments sections and newsboards. This is invaluable information that will help you discover who your audience is and how to hit their mark.

There will be a lot of criticism—from hockey traditionalists, from fans, from the media, and most especially from your colleagues within the hockey analytics community itself. No matter what you do, there will always be about five to ten critics for every supporter. Always remember that the only way negative people can be right about you is if you quit.

Finally, Lenox's closing advice was to write every day and the ideas will surface. There is absolutely no substitute for hard work, and you'll need a genuine passion for this field in order to stick with it.

Although it was meant for a completely unrelated field, Lenox's advice got me a long way in hockey analytics. Since then, I've been involved in a lot of hiring decisions, and I've picked up the following more practical and specific pieces of advice directly from those in front offices:

1. CREATE SOMETHING
Whether you write a paper or an online article, hold a conference, invent a statistic, build a website, found a company, participate in a manual tracking project, or start a podcast, it helps to create a name and an identity for yourself.

2. NETWORK AND BUILD UP YOUR CONTACTS
Everybody can teach you something valuable if you have an open mind and listen. With regard to jobs, most of them are not posted publicly, and those that are generally receive between 500 and 1,000 responses, depending on the market. Either way, those who have contacts are those who hear about these opportunities first and get on the short lists.

3. GET SOME WORK EXPERIENCE
Contact the junior, college, and minor-league teams in your area and offer your assistance. Think outside the box, and contact player agencies and third-party consulting companies. Even if most of them ignore you, or even if it's part-time or unpaid work, you only need one response to get started.

4. TAKE YOURSELF TO THE NEXT LEVEL
Whether you're into manual tracking, programming, or statistical analysis, take the time to master your craft. That could mean studying textbooks, using software and other technology, or getting lots of practice in order to complete a wider variety of tasks or to complete them faster and with greater accuracy (just like an NHL player would).

5. BUILD A PORTFOLIO
When you get that meeting and/or interview, it helps to be prepared with examples of your work, even if it's just in one small area of the game. Make it something practical and memorable, and tailor it to the organization in question, if possible.

6. GET YOUR FINANCIAL HOUSE IN ORDER
Quite frankly, it's shocking how little most of these jobs pay compared to similar work in other fields. This is a by-product of how badly so many people want to work in hockey and how many organizations aren't yet properly budgeted for analytics. It would be highly unfortunate to pass on a great opportunity because of debts, a high cost of living, or the inability to relocate.

7. DON'T BE AN ASS ON SOCIAL MEDIA
I have personally seen some golden opportunities get flushed down the toilet because someone was acting highly unprofessionally and disrespectfully on social media, usually by trashing players, front

offices, journalists, or fellow statisticians. Most of the time they didn't even realize that they had blown a great opportunity. And for goodness sake, do not write about sex, religion, money, or politics, the last of which appears to be the hardest to resist!

8. CONFIDENTIALITY IS KEY

The NHL is understandably a very competitive and secretive league, so discretion is imperative. It can be hard to resist shouting it from the rooftops when you've done something amazing, but most teams simply can't risk working with anyone who might leak information. It's important to have (and keep) a good reputation in a relatively small field where most people keep in semi-regular contact.

9. SHOW, DON'T TELL

Whenever you see a statistician fail to make a point about Corsi or PDO, it's probably because they're telling others how it works rather than showing them. Part of the appeal of player usage charts is that we're not telling the reader about zone start percentages, of quality of competition, or usage, or Corsi—we're showing them. Likewise, don't go to a meeting or interview to tell someone what you can do for their organization; go prepared to show them. That might mean video or a demonstration or a chart or some tables or specific recommendations—who knows? But find specific ways to show them how you can help.

10. BE PREPARED FOR SACRIFICE

Remember: the passions of even the fortunate few who were offered front-office positions went unrewarded for the longest time. And even when they landed their first great opportunity, they were asked to shut down their life's work, quit their (generally) lucrative day jobs, and toil for an average of 66 hours per week in potentially a very limited corner of the organization.[123] That's a sacrifice that only the truly devoted can make.

---

123. Michael Lopez, "What's It Like to Work in Sports Analytics?," *Stats by Lopez* (blog), January 15, 2017, https://statsbylopez.com/2017/01/15/whats-it-like-to-work-in-sports-analytics/.

Despite the premium placed on discretion, my involvement in more than a few of the recent hires has been a rather poorly kept secret. Without confirming any of the specifics, I can explain how this came to be.

It started about four years ago. It might surprise you that NHL teams had difficulty finding and identifying good analysts, but that was the case at the time, and to an extent remains so today. It's not like there's a minor-league system for statisticians or a lot of schools or recruiting firms that specialize in hockey analytics.

As one of the more visible people in this field, I started getting a lot of phone calls from various front offices to get opinions on certain individuals or to see if I knew anybody who could help them. At first, I put them in touch with people I knew, including those whose work had been featured in the original *Hockey Abstract*, but I ran out of friends pretty quickly. Those who weren't snapped up in the 2014 summer of analytics were certainly gone by the end of 2015. Plus, there were so many talented new people arriving to the scene that I couldn't even be confident that I was actually presenting these organizations all of their best options.

That's why I opened things up in February 2016 and invited everybody who might be looking for a future in hockey analytics to contact me and to let me know what they're looking for.[124] Then, I could pass along the right opportunities to the right people and give front offices even more options from which to choose.

As you would expect, the response was a little overwhelming, and hundreds of people reached out to me. Some people were just looking for a little bit of advice and encouragement, or maybe just a little part-time hobby work on the side, but others were deeply passionate about full-time careers in the sport.

It took me nine months, but after taking up to a dozen calls per week, I managed to follow up with all the initial respondents and helped a lot more people get started. Several dozen got audiences

---

124. Rob Vollman, "Do You Want an NHL Job in Hockey Analytics," Hockey Abstract, February 9, 2016, http://www.hockeyabstract.com/thoughts/dowantannhljobinhockeyanalytics.

with NHL front offices, and several of them got full-time jobs this season. Of course, during the time I wrote this book, my backlog swelled to an even greater size than before. Once more into the breach!

People ask me why I do this, but it was a highly rewarding and educational experience for me. I was energized from meeting so many bright and passionate young people, and I learned about some of the exciting projects they're working on. I also got the opportunity to learn a lot about how teams are building their analytics departments, what kinds of opportunities are available for those in our field, what skills teams are looking for, and what people have done to get ahead, all of which I've shared here.

Ultimately, I think our careers in hockey analytics are whatever we create for ourselves. When I started, there were no hockey analytics books, no hockey stat websites, and no statisticians discussing our sport in newspapers, online, on TV, or on radio. There were virtually no statisticians working in front offices, no hockey analytics conferences, no visualizations, and very few stats beyond goals, assists, and plus/minus. We had to create all of these opportunities from scratch.

If anything, front offices and other hockey organizations have their ears and eyes open for us to tell them what opportunities should exist and how to create them. If you wait until an opportunity exists before you go for it, then you might just miss it.

QUESTIONS AND ANSWERS

One of my favourite applications of hockey analytics is to answer questions. These are often topical questions, like who won the Shea Weber for P.K. Subban trade, but others are timeless enough to include here.

Usually, the demonstration of how stats can be used is more interesting than the answers themselves. However, that all depends on the question. In this book, we'll look at the value of a penalty shot versus a power play, when teams should pull the goalie, if Ovechkin will ever catch Gretzky, how the NHL can boost scoring, and what might be the key individual player statistic. Let's begin.

Which Is Better, a Penalty Shot or a Power Play?

This is more of a theoretical question, since neither team has any control over whether an infraction will be a penalty shot or a penalty, but it is an interesting question nonetheless.

A penalty shot is obviously better than a 5-on-4 power play. Going all the way back to the 1934–35 season, NHL players have scored on 565 of 1,561 penalty shots, which is 36.2%.[125] Going back to just the 2005–06 season, it's 246 goals in 745 shots, which is 33.0%. That's far better than a 5-on-4 power play, which has ranged between 17.3% and 19.1% over this time span and carries the risk of a short-handed

---

125. Shootout data from the National Hockey League, accessed August 1, 2017, http://www.nhl.com.