

Katerina Wu

November 1-3, 2019

### **Adventures in Economics Report**

This past weekend I had the exciting opportunity to attend the annual Carnegie Mellon Sports Analytics Conference. Although it's a relatively new conference (this was only Carnegie Mellon's third year hosting) compared to similar symposiums within the field, like MIT's Sloan Sports Analytics Conference, it already attracts a deeply dedicated audience. Pittsburgh has become a flourishing center for the sports analytics community in recent years, and the results were noticeably reflected in the quality of the speakers and wide variety of talent present at the conference.

I was initially drawn to attend CMSAC as part of my research for my senior honors thesis. I was studying an iteration of the principal-agent problem in labor economics. My thesis focused on the effects of long-term guaranteed contracts on player performance, specifically within the National Hockey League. After hearing influential figures in the sports community like Sam Ventura of the NHL's Pittsburgh Penguins were going to be present, I immediately reached out to the director of the conference, Professor Rebecca Nugent, to see if there were accommodations available so I could speak with some of the presenters face-to-face and gain their insight to the topic of my thesis from another perspective. She reached back with an affirmative, and also encouraged me to submit an abstract for their reproducible research competition. I was now signed up to attend and to present during the poster session.

I left UNC for the airport early Friday morning despite my afternoon flight; part of my luggage included a poster tube for my presentation, and I was worried that there might be

complications regarding the TSA check-in. However, the entire process went smoothly, and I soon found myself landing safely in Pittsburgh around 4PM. I took the bus which ended up being an idyllic one hour ride to the heart of the Carnegie Mellon University campus. Once there, I met up with a friend from the University of California at Fullerton who also was interested in sports analytics. This was both of our first times attending an event like this, and we weren't sure what to expect. Together, we headed towards the conference's first event: the football analytics workshop.

The CMSAC football analytics workshop was a three-hour event led by Ron Yurko. The first hour dedicated to introducing attendees to reading, wrangling, and visualizing publicly available NFL data with the R statistical programming language. The third hour of the workshop covered the basics of using R to generate ELO ratings for NFL teams, a popular rating system featured on websites such as FiveThirtyEight. This was particularly interesting to me, because my capstone project focused on a similar topic of generating ELO rankings for international ice hockey leagues. The middle hour of the workshop featured the keynote speaker, Michael Lopez, who discussed his work as the Director of Data and Analytics for the National Football League. No prior programming experience was required, and all the instructions provided were clear and straightforward. One of the goals of CMSAC was to create a welcoming environment for people who were just getting started in sports analytics, and they certainly delivered on that front.

On Saturday we woke up with a fully packed agenda ahead of us; the entire day was filled with presenters and workshops to attend. The first presentation I attended was about unexpectedly about history. Professor Christopher Phillip of CMU's history department educated us with indispensable information regarding the history of data collection. No matter what field

the research is being conducted in, he emphasized, it is important to consider where the data is coming from and what inherent biases are present. For example, in MLB data, there used to be measurements for how “gentlemanly” a batter was, which seem ridiculous to us now but was common practice at the time. In my ECON 570 class with Professor Hill, this topic was briefly mentioned, and I appreciated seeing this connection in different fields as well.

Following the morning session of presentations, there was a brief interlude where the attendees went out for coffee and lunch. In between the breaks we mingled with other students and representatives from various sports-related correspondents, such as analysts from the Cleveland Indians and agents from Big League Analytics. I was even able to find another sports economist in attendance from Syracuse University! We had a scintillating conversation on his research, which focused on the effect of online betting odds on player performance.

During one such coffee break, I met the man I mentioned before in my report: Sam Ventura. It’s a rather surreal experience to meet someone that you discuss within your thesis. Sam is the Director of Hockey Research for the Pittsburgh Penguins of the NHL. He spearheaded many endeavours for more accurate predictive models of player performance and stability, including the preeminent “WAR-on-ice” movement. I utilize and reference a large quantity of the logic from his papers in the literature review portion of my thesis, as well as for the controls within my empirical model. When I spoke to him about the topic of my thesis, he readily provided me suggestions and feedback for how to implement them. He even offered to send me some proprietary data on player tracking information that might assist in better measures for my response variable, which is going to be invaluable to the quality of my analysis.

After that brief break, we all reconvened for the afternoon session of the presentations. Many students presented during this portion, and judges roamed around the auditorium to evaluate the quality of the poster presentations. To my extreme surprise, I was actually later announced to be the winner of the Poster Presentation Competition for undergraduate students! Considering the quality of the work around me and the sheer talent and creativity of the other posters, for my work to be considered on the same level as theirs was such an extreme honor.

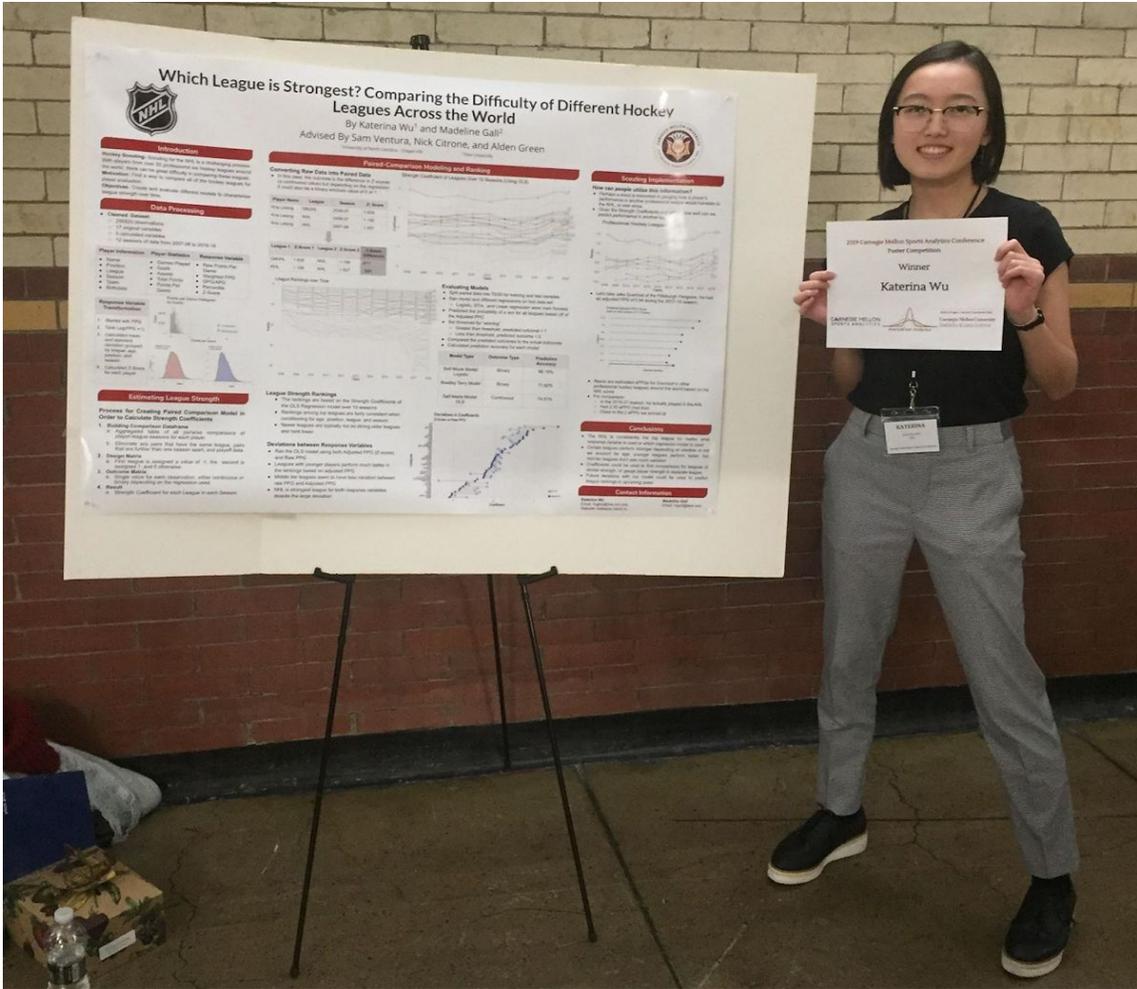
Even more shocking was I to learn that so many people found my research and ideas to be thoughtful and well-explained. Multiple people came up to me after my presentation to congratulate me and commend me for my work, even academic idols that I looked up to online. Throughout the entire conference, various people struck up conversations with me about my future goals for research, and what new ideas I was currently working on for my thesis, even if it was more economics-oriented than typical sports analytics. I was blown away by the enthusiasm and support. Sports analytics is a heavily male-dominated field so it always felt intimidating to me when I first started gaining interest, especially as a woman of color. However, everyone I spoke to was incredibly uplifting and left me feeling affirmed for my involvement in such a community.

One such conversation that left an indelible mark on me was the brief chat I had with Dani Chu. Dani was recently hired to work in the front office of the NHL's newest team located in Seattle. He described how he first got started with sports management and business, but slowly started expanding his research into other fields, and how he amalgamated his various skills into being a hockey analyst. He then encouraged me to apply for the quantitative analyst

position for the Seattle team. Although it's been a latent dream of mine to work for a hockey team, I never thought it was actually a feasible goal until now.

A similar interaction happened with Asmae Toumi. She is the editor-in-chief of Hockey Graphs, a site dedicated to visualizing and analyzing hockey statistics. It consists of writings from prolific researchers and well known analysts within the hockey world, who all produce high-calibre content. After my presentation, Asmae came down to introduce herself and personally invited me to do a write-up of my work for her website. These two conversations really caused a paradigm shift within my perception of my own abilities, and made me realize that these aspirations that seemed like mere fantasies before are well within my reach.

Ultimately my time in Pittsburgh came to an end. I took an Uber back to the Pittsburgh International Airport bright and early Sunday morning at 5AM for my 7AM flight. I learned so much throughout the entire process (from other people, and about myself) and also came away with a newfound fervour for research. Of course, this wonderful opportunity would not have been possible without the support of my professors in the Economics Department here, or the generous donations from the sponsors of the Adventures in Economics fund, so many thanks are due! I hope future recipients of the award will also encounter a great adventure like mine.



# Which League is Strongest? Comparing the Difficulty of Different Hockey Leagues Across the World

By Katerina Wu<sup>1</sup> and Madeline Gall<sup>2</sup>  
Advised By Sam Ventura, Nick Citrone, and Alden Green  
<sup>1</sup>University of Pennsylvania, <sup>2</sup>Carnegie Mellon University



### Introduction

Which league is the strongest? According to the NHL, it is the National Hockey League. However, this is not necessarily the case. In this paper, we compare the difficulty of different hockey leagues across the world. We use a variety of metrics to evaluate the strength of each league, including the number of teams, the quality of the players, and the level of competition. Our findings suggest that the NHL is not necessarily the strongest league, and that other leagues, such as the KHL and the AHL, may be more difficult to play in.

### Data Processing

• Collected data from NHL, KHL, AHL, and other leagues.  
• Cleaned and processed the data.  
• Analyzed the data using various statistical methods.

### League Strength Ratings

• The NHL is the strongest league, followed by the KHL and the AHL.  
• The difficulty of each league is measured by a variety of metrics, including the number of teams, the quality of the players, and the level of competition.

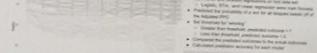
### Comparing Raw Data into Player Data

• League Strength Rating (LSR) is the sum of the following metrics:  
• Number of teams  
• Average number of goals per game  
• Average number of assists per game  
• Average number of points per game  
• Average number of minutes per game

League	LSR
NHL	1.00
KHL	0.85
AHL	0.70

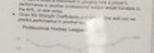
### League Strength vs. League Type

• The NHL is the strongest league, followed by the KHL and the AHL.  
• The difficulty of each league is measured by a variety of metrics, including the number of teams, the quality of the players, and the level of competition.



### Scoring Inflation

• The NHL has a higher number of goals per game than other leagues.  
• This is due to a variety of factors, including the quality of the players and the level of competition.



### Conclusion

• The NHL is the strongest league, followed by the KHL and the AHL.  
• The difficulty of each league is measured by a variety of metrics, including the number of teams, the quality of the players, and the level of competition.

### References

• NHL.com  
• KHL.com  
• AHL.com

2018 Carnegie Mellon Sports Analytics Conference  
Student Competition  
Winner  
Katerina Wu

KATERINA WU