Rachel Wantuch

DSC 530 – Metzger

**Final Project**

# Statistical Questions/Hypothesis

What makes a hockey player great? How do the legendary players become legendary? Is there a combination of attributes that can predict how long a career will be?

# Outcome:

The exploratory data analysis (EDA) shed light on a few patterns in the data to help answer the questions above. The histograms all showed the same pattern of right skewed histograms. This makes sense in the context of examining sports because everyone will have played at least one game, for example, but fewer and fewer people will have played a greater amount of games.

# What do you feel was missed during the analysis?

I’m not complaining because I don’t have a lot of free time to start with but my analysis could have benefitted from doing a few more regressions and maybe doing a more thorough analysis using the PMF and CDF aspects. I felt that I accomplished a very cursory overview of the data but it wasn’t thorough enough to answer the statistical questions in depth.

Overall I feel like I needed to hone in my question a little more too. It was a bit too general to gain any specific insight to what makes a successful NHL player.

# Were there any variables you felt could have helped the analysis?

There was one variable that I thought of towards the end that could have been used as a benchmark for a “successful career”; career earnings.

# Were there any assumptions made you felt were incorrect?

I was making an assumption that penalty minutes was a beneficial aspect to consider when in fact the game isn’t as physical as it was in the past. Also I examined the career stats and perhaps changing those to averages over the number of games they played may have been a better read on how effective that player was.

# What challenges did you face, what did you not fully understand?

The first challenge was having enough time in my own life. It is what it is. The second challenge was understanding the outputs for the regressions better. I just got too rushed towards the end and went with the simpler OLS regression for one dependent variable and one independent variable because I knew what the results meant off hand. I feel like I also need a better understanding of choosing good variables for the PMF and CDF aspects to gain more insight.