**What is a “Good Player” and What makes a successful season?**

* Hockey is a team sport and it is hard to identify individual contributions to the overall outcome and what makes a championship team.
* This project will examine what statistics or measures of “Team Performance” are most indicative of a successful season (Making it to the conference finals). In addition, after looking at what predicts a successful season, I will look at classifying players based on their salaries to determine the efficiencies in building a good team.
* This impact could help understand some of the more nuanced forms of player analysis. And if I am super successful and the stars align I will become the general manager of the Rangers.
* The analysis will be based on two very interesting machine learning methods (and maybe more that I have not learned yet). First will be to see what statistics or predictors are the strongest for predicting a strong season using different regression models, most likely a Logit model for “successful” or not. Then utilizing a classification model, possibly something more robust than a Neighbors classified like a SGD Classification which may be useful given the massive amount of data I have access to.

**Datasets**

* I have access to individual game data along with season data for each player and team in the NHL from 2000 until the current date thanks to War-On-Ice.com
* There is access to a web scraper via github that processes newly created NHL data, although I have yet to use it.
* General Fanager.com provides an easily scraped table of team contract and salary data that is updated daily.
* HockeyReference.com has data for each team historically going back since the league was created and has a simple to use interface to gather data.
* Three other websites provide ancillary information that may prove useful

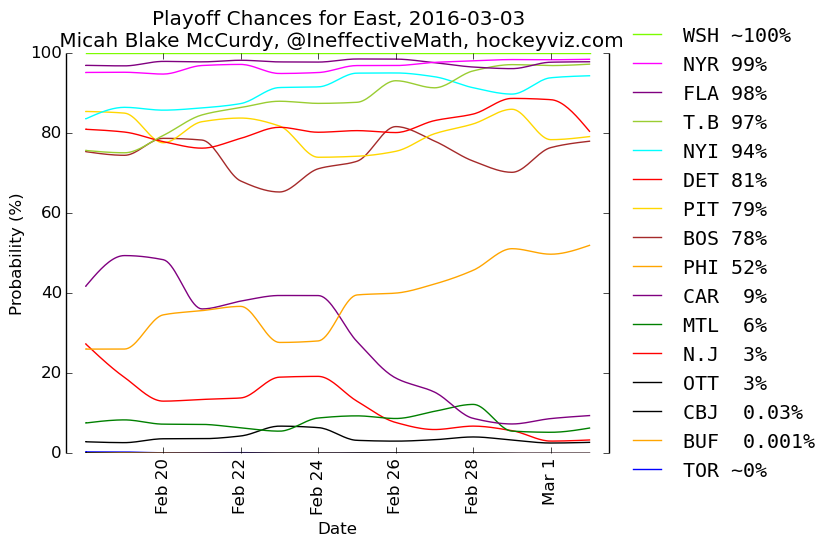
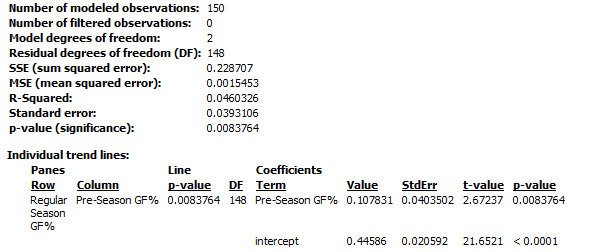
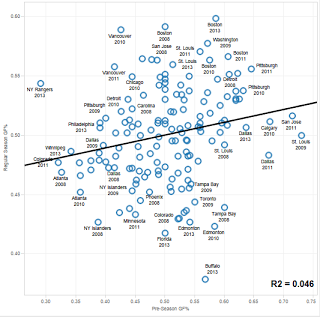
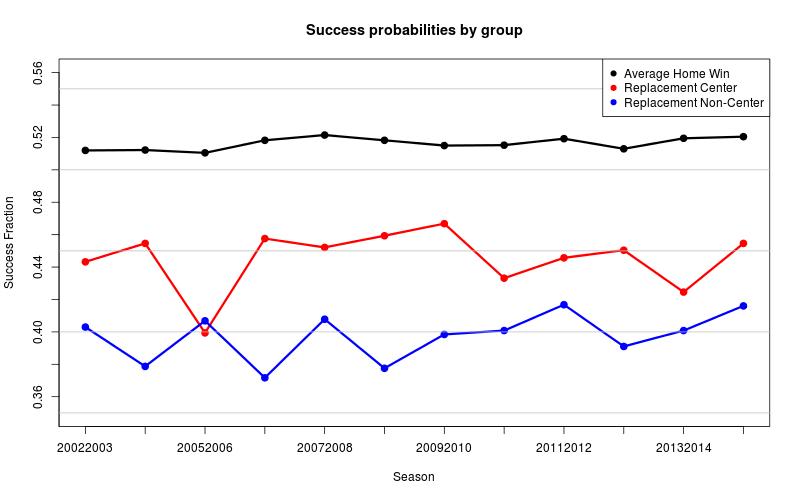
**Domain knowledge**

* I have played hockey for most of my life so I have an innate knowledge of the subject. Furthermore, I follow the sport on a daily basis as well as read the latest in analytics developments so I’m familiar with how people are applying data science in other ways.
* There are many other people who are working in this area and are very happy to share their work.

**Project Concerns**

* My biggest concern with my project is becoming lost in the data and not knowing what model works best.
  + I have access to a lot of data and many features that could be relevant; finding ones that are the most effective will be time intensive and computationally heavy.
* I will most assuredly need guidance in determining the proper model and right machine learning technique to progress with my project.
* Hockey is generally viewed as a game of “random bounces” that normalize over the course of a season, which is appealing, but trying to find the right test sets is going to be hard.
  + Is each test set going to be each team, each season, each team by season or possibly by player and outcome.

**Outcomes**

* The desired outcome for me would be to find a predictive model using 4 or fewer features related to player performance that can identify likelihood of season success.
* **Similar Projects**
* Logit Model for Probability of Making Playoffs
* Pre-season impact on final standings:
*  
* Calculating “Wins Above Replacement” and anticipated “Success”
* 
* 
* **SOURCES:**
* [**http://blog.war-on-ice.com/sharing-is-caring/**](http://blog.war-on-ice.com/sharing-is-caring/)
* [**http://war-on-ice.com/playertable.html**](http://war-on-ice.com/playertable.html)
* [**http://www.hockey-reference.com/teams/NYR/1999.html#skaters::none**](http://www.hockey-reference.com/teams/NYR/1999.html#skaters::none)
  + If I am unable to find a good model (what is good to be determined as I progress) I may look at non-player determined variables like team salary, weather, time of games, etc.
* Given what I’ve seen from peers on the internet in the data analytics community what people are looking for is a way to understand what happens in a game to a quantitative form. For example, “By having a higher number of assists, a team is more likely to make the conference finals.” Then using that result to look at individual players and see how they contribute to the team.
* I think my project won’t be a bust because if I am unable to find anything that can adequately predict “success” then those in the hockey community who say “analytics and big data are overrated” will have more ammunition.