Simulating The 2024 NHL Playoffs With Luck

Bella Salter

2024-04-16

In the hockey community, there are various superstitions regarding the impact of rankings on playoff performance. Some claim the President's Trophy brings both bad luck along with its glory, but others disregard superstition, firmly believing in rankings as a predictor of playoff success. Do regular season rankings truly indicate a team's luck in the playoffs?

During the regular season, teams earn points depending on their performance in every game. If they win, they earn two points, and zero for a loss in regulation. Alternatively, if they lose in overtime, they still earn one point. Then, the team with the most points is ranked first, and so on. The top three teams in each of the four conferences are selected for the playoffs, and then the two teams with the best records in each division are selected as well. Typically, the playoffs end up consisting of the sixteen best-ranked teams in the league, but of course there have been a few oddities over the years. In the playoffs, teams play seven games against their opponent to decide who moves on to the next round. We can first attempt to analyze the effect of regular season rank by creating a linear model to predict playoff rankings.

```
simple_model <- lm(hd$PORK~hd$RK)
summary(simple_model)
```

```
##
## Call:
## lm(formula = hd$PORK ~ hd$RK)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -20.0101  -3.9012   0.0072   4.7177  13.1533
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.08447    1.03719   3.938 0.000124 ***
## hd$RK        1.05445    0.05618  18.768  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.328 on 154 degrees of freedom
## Multiple R-squared:  0.6958, Adjusted R-squared:  0.6938
## F-statistic: 352.2 on 1 and 154 DF,  p-value: < 2.2e-16
```

Here, we can see that a linear model to predict playoff ranking based only on regular season rank has a relatively high R-squared value for training on one aspect of a team. This means that regular season rank could be a predictor of success, but intuitively, the better teams should be ranked higher, since they should win most of their games in the regular season. Additionally, it is widely known that playoff hockey is played very differently than regular season hockey. Not only do the referees adopt a laissez-faire mentality, but higher stakes come playoff season take a toll on players' mentalities and performance. Let's take a look at some trends of the top two playoff teams for the past few years to compare some regular season statistics.

```
## [1] "Regular Season Rank: mean = 7.4, standard deviation = 5.89161"

## [1] "Goals For / Game: mean = 3.28, standard deviation = 0.262128"

## [1] "Goals Against / Game: mean = 2.805, standard deviation = 0.200458"
```

Here, we can see that the average regular season rank of a top-2 team in the past five years has been 7.4, with quite a large standard deviation. This implies that despite rankings' supposed indication of a quality team, highly ranked teams may not necessarily perform well in the postseason. Let's visualize the typical regular season ranking of the past five Stanley Cup winners.
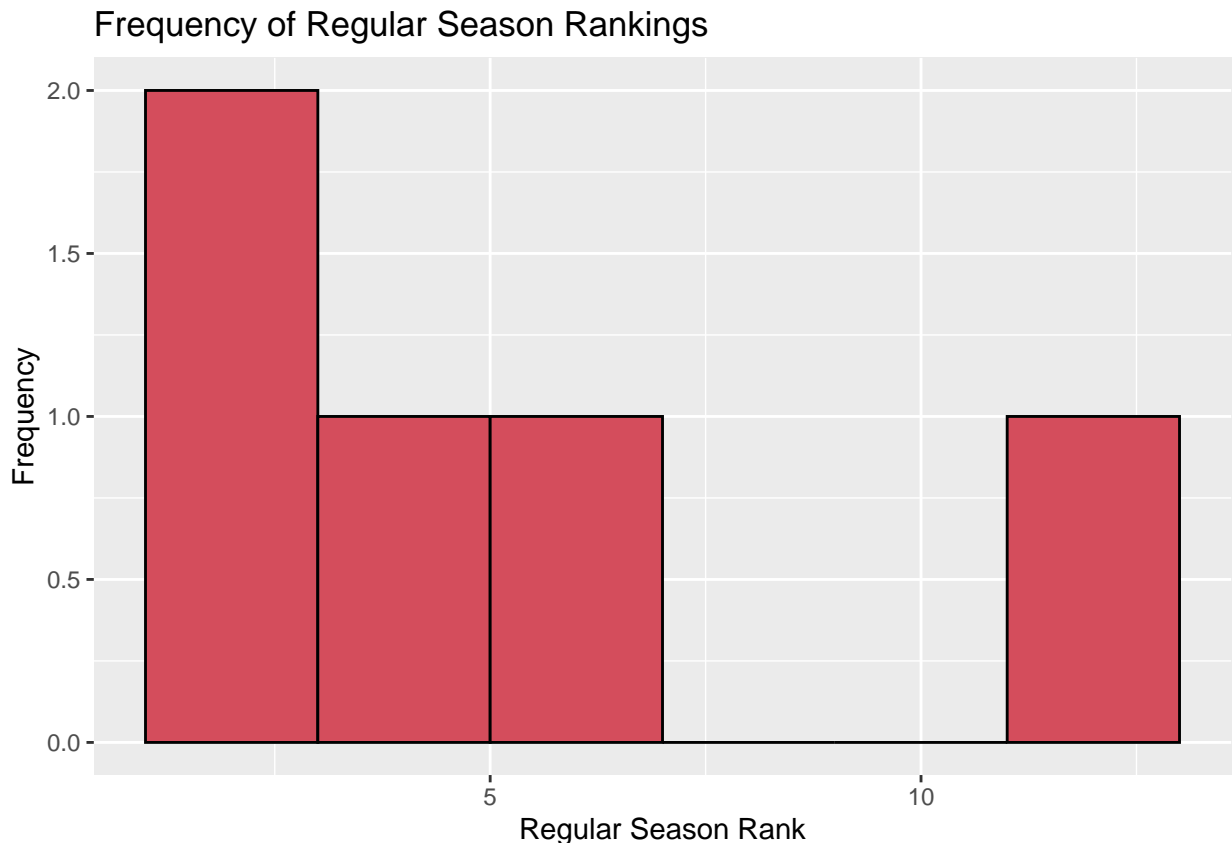
Here, it can be seen that it is more common that a team ranking lower than first wins the Cup. In fact, only eight teams have won the Presidents' Trophy(awarded to the first-ranked team) and the Stanley Cup in the same year. Therefore, when simulating playoff games, perhaps rank could be used as a measure of "luck" for a team. To analyze the effect of regular season ranking on performance during each round of the playoffs, I'll plot the average playoff ranking for each regular season ranking.

Average Ranking Changes in Playoffs
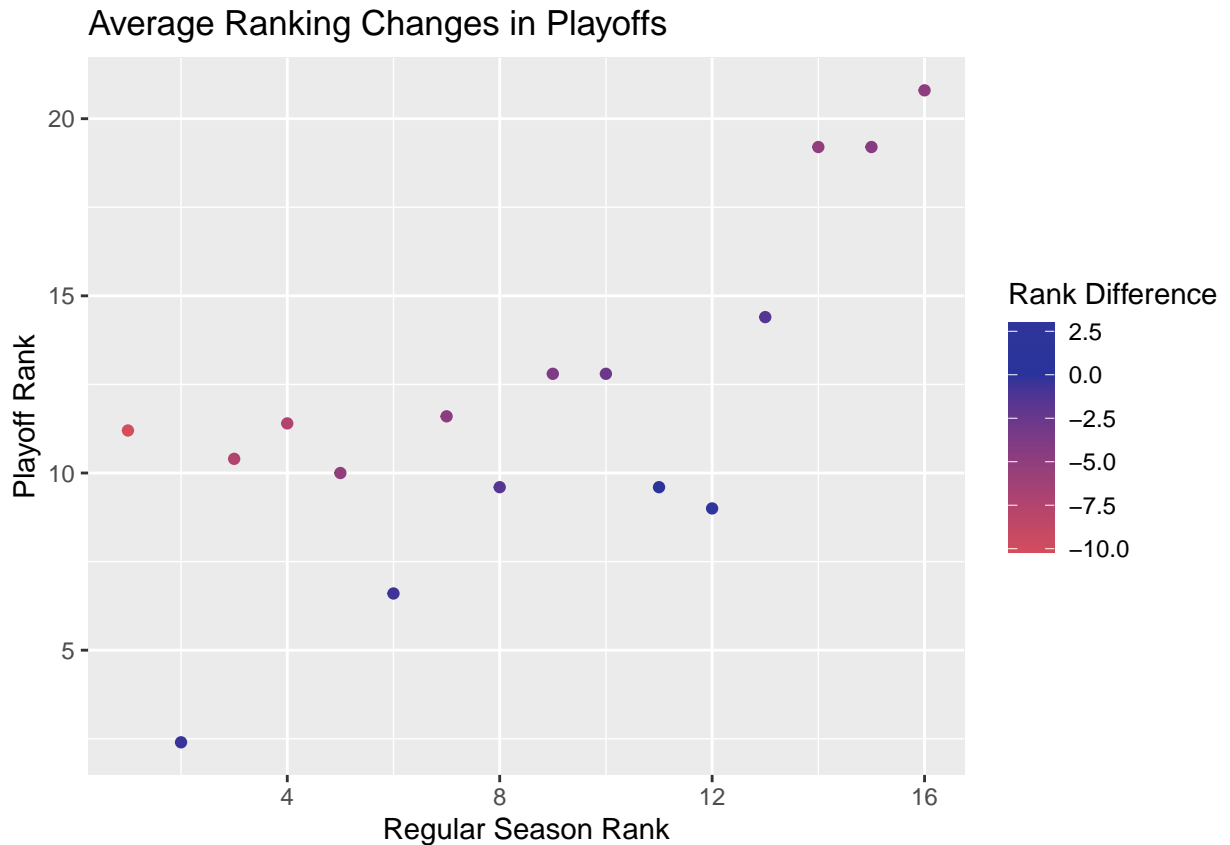
Here, we can see teams ranked second, fifth, and twelfth perform better in the playoffs than expected. On the other hand, first and third teams play abnormally worse than their ranks would suggest. Since the rankings themselves do not serve as a full indication of the playoff outcomes, perhaps the games could be simulated using the teams' average goals for/against per game, with rank as a factor to represent luck. In order to assess the quality of the simulation, I will be using data from the 2022 season from hockey-reference.com. Before simulating games, let's take a look to determine the distribution of goals in each game for each playoff team.

Frequency of Goals Scored
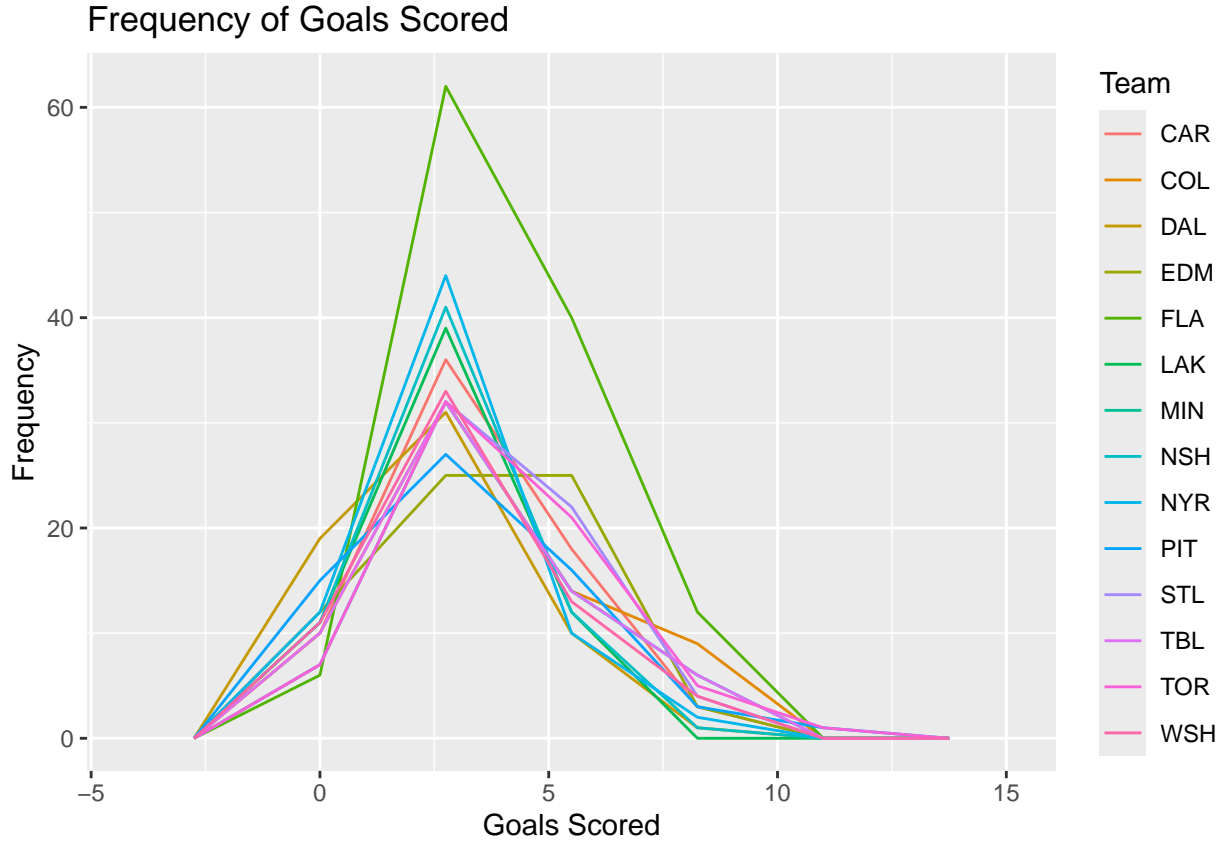
This plot looks roughly normal, so I will assume that goals scored are distributed normally in each game. This is also common practice in sports such as soccer. Since goals scored for a team is distributed normally, I also assumed that goals scored against a certain team is distributed normally. With these assumptions, a game can be simulated by sampling from normal distributions to determine if a goal was scored by either team in each minute of a game, tallying these goals for each game, and repeating so that we can use the law of large numbers for our approximation of the winner. To determine the distribution of the goals scored for each team to sample from, I found the mean($\bar{X}$) and standard deviation(S_X) of their goals for and against during the season. This is because these are the unbiased estimators of expected value and variance, which are the parameters for normal distributions:

$$\bar{X}_N = \frac{1}{N}\sum_{k=1}^{N} X_k \approx \mu_X$$

$$S_X{}^2 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(X_k - \mu_X)^2} \approx \sigma_X{}^2$$

For example, for the Vancouver Canucks this year, this would be calculated in R as follows:

```
canucks <- get_all_goals(info24, "VAN")
sprintf("Mean = %g, Standard Deviation = %g", mean(canucks$gf), sd(canucks$gf))
```

```
## [1] "Mean = 3.47761, Standard Deviation = 1.94903"
```

Now, for each team we assume that the goals scored for the team in each team is distributed normally

4

according to the approximated parameters:

$$\mathbb{P}(X \leq x) = \frac{1}{S_X \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - \bar{X}_N}{S_X}\right)^2\right)$$

Finally, to run the simulation, we calculate these statistics above for the goals for and against each team per minute. Then, we sample from these normal distributions for 60 minutes, determining a winner based on who has scored more goals. Finally, we repeat this 10,000 times to determine our winners from the first round. In 2022, the first round matchups were:

```
##    Team 1     Team 2
## 1    CGY vs.     DAL
## 2    COL vs.     NSH
## 3    FLA vs.     WSH
## 4    NYR vs.     PIT
## 5    EDM vs.     LAK
## 6    MIN vs.     STL
## 7    TBL vs.     TOR
## 8    BOS vs.     CAR
```

To simulate luck, I chose to divide 16 minus the team's rank divided by a factor of 320000. I then added the luck statistic to the average for the distributions of each team. Running the simulation 10,000 times gives us the first-round winners as:

```
## [1] "CGY"
## [1] "COL"
## [1] "FLA"
## [1] "NYR"
## [1] "EDM"
## [1] "MIN"
## [1] "TOR"
## [1] "CAR"
```

Now, we can do the matchups as the NHL would and continue to find our second-round winners.

```
team1 <- c("EDM", "COL", "CAR", "TBL")
team2 <- c("CGY", "MIN", "PIT", "FLA")
rd2 <- data.frame(Home=team1, Visitor=team2)
sim_round(rk_info, game_info, rd2, 1)
```

```
## [1] "CGY"
## [1] "COL"
## [1] "CAR"
## [1] "TBL"
```

We repeat this twice more to find our 2022 champion:

```
## [1] "Top 2 Teams:"
```

```
## [1] "COL"
## [1] "TBL"
```

```
## [1] "Stanley Cup Champion:"
```

```
## [1] "COL"
```

Therefore, we would find that our Stanley Cup Champion would be the Colorado Avalanche. This did in fact occur in 2022, so this is a promising outcome. In fact, the model was only incorrect about Calgary and Carolina winning in the second round, and Minnesota and Toronto in the first. However, it could be that the "luck" variable just so happened to accurately describe the 2022 playoffs. How would this play out in 2024? As of April 19, we have the matchups for the playoffs and can simulate them the same way.

```
## [1] "VAN"
## [1] "CAR"
## [1] "BOS"
## [1] "FLA"
## [1] "NYR"
## [1] "COL"
## [1] "EDM"
## [1] "DAL"
```

Using these first round winners, we can then use the same method of simulation to end up with our final two players and winner.

```
## [1] "DAL"
## [1] "VAN"
## [1] "BOS"
## [1] "CAR"
```

```
## [1] "Top 2 Teams:"
```

```
## [1] "DAL"
## [1] "BOS"
```

```
## [1] "Stanley Cup Champion:"
```

```
## [1] "DAL"
```

According to the simulation above, this would mean that the Dallas Stars will win the cup in 2024. This is a somewhat promising result, as many sports writers have selected the Stars as their favorite to win this year. And, of course, they have one of the highest goal differentials in the league.

However, of course, this comes with errors. First, many generalizations had to be made to account for luck. Perhaps rank is not the only indication; average Stanley Cup wins of the players could also be a factor. Furthermore, this is a very simple model of a hockey game. Future work could improve upon this model by more accurately describing the distribution of the goals scored, or by factoring in statistics such as amount of takeaways and giveaways for each team. All in all, time will tell how accurate this simulation is when applied to 2024 data.