An Exploration of NHL Coaching Statistics

Introduction

The datasets used for this project were derived from the Professional Hockey Database from Kaggle (https://www.kaggle.com/open-source-sports/professional-hockey-database (https://www.kaggle.com/open-source-sports/professional-hockey-database)). This data, originally sourced from The Hockey Database from Open Source Sports, is a collection of historical statistics from men's professional hockey teams in North America from 1917 until 2011. Although there were 20 datasets within the source, only 6 were utilized (masters, coaches, awardCoaches, teams, teamsPost, and abbrev). Masters includes the names and biographical information for all players and coaches involved in the North American Hockey, coaches includes coaching statistics, awardCoaches includes coaches awards, trophies, and postseason all-star teams, teams includes team regular season statistics, teamsPost includes team post-season statistics, and abbrev includes abbreviations used in Teams and SeriesPost tables.

Using these datasets, we were able to answer the following research questions for the NHL:
* How does the team performance change after the coach wins the Jack Adams award?
* How does the team performance change after winning a Stanley Cup?
* What coach coached the same team for the longest number of years?
* What coach won the most Stanley Cups?
* What coach had the best full regular season (in terms of number of points per season)?
* Is there any correlation between regular season and post-season performance?
* Are there any coaching tactics that consistently result in a good record? High PKC / GA (Defensively oriented) or High PPG / GF (Offensively oriented)? (PKC = Penalty Kill Chances, GA = goals against, PPG = power play goals, GF = goals for)

It is important to highlight that the points system in hockey is as follows: a win is worth 2 points, an overtime loss is worth 1 point, and a straight loss is worth 0 points. Furthermore, the modern NHL playoff (based off of the team with the most amount of points in their division at the end of the regular season, plus wild cards from each conference) structure started in 1988. Lastly, in 2004, the NHL experienced a lockout and had no games as a result.

Data

Question 1

How does the team performance change after the coach wins the Jack Adams award?

answer
We are looking at the teams that had a coach win the Jack Adams during the 5 year period from 1995 to 2000, and how well that team did for the following five years. It appears that no general trend, other than the fact that almost every team does worse the succeeding season, is present for these teams. We are unable to conclude anything further about the performance after a coach wins the Jack Adams.

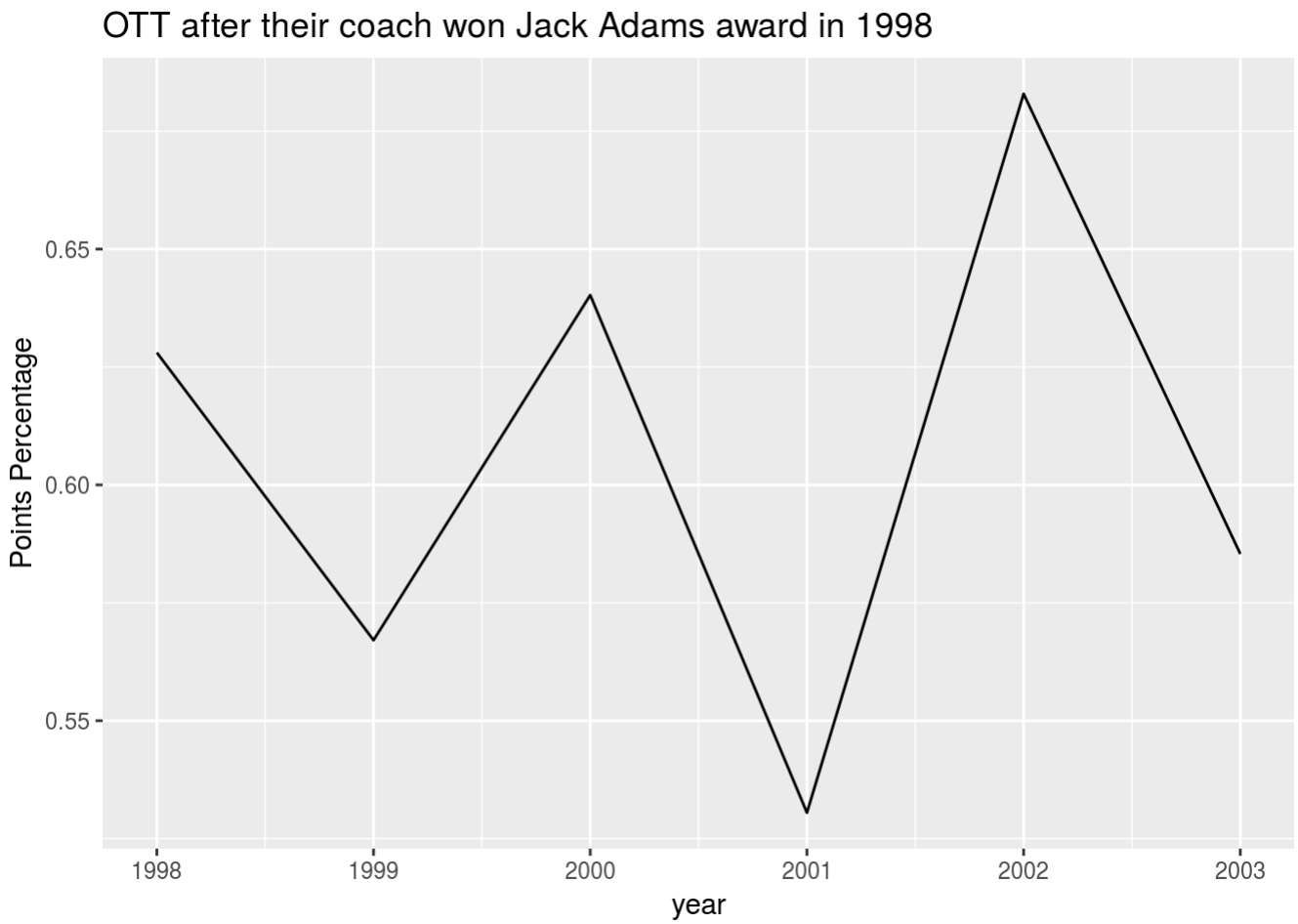
```
## # A tibble: 6 x 9
##   coachID     year tmID      g     w     l     t award      won_award
##   <chr>      <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <chr>      <lgl>
## 1 bowmasc01c  1995 DET      82    62    13     7 Jack Adams TRUE
## 2 nolante01c  1996 BUF      82    40    30    12 Jack Adams TRUE
## 3 burnspa01c  1997 BOS      82    39    30    13 Jack Adams TRUE
## 4 martija02c  1998 OTT      82    44    23    15 Jack Adams TRUE
## 5 quennjo01c  1999 STL      82    51    20    11 Jack Adams TRUE
## 6 barbebi01c  2000 PHI      54    31    16     7 Jack Adams TRUE
```

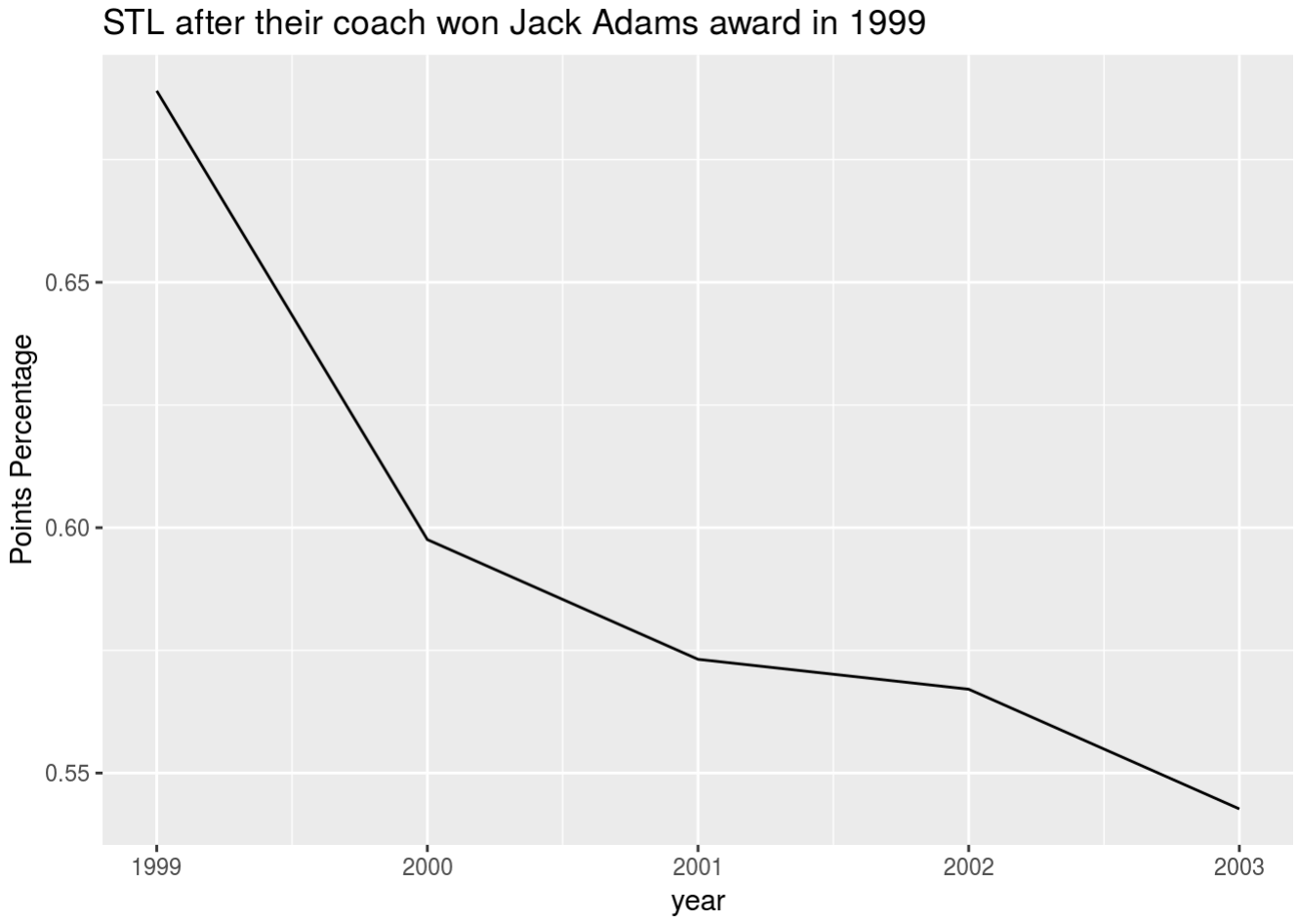
DET after their coach won Jack Adams award in 1995

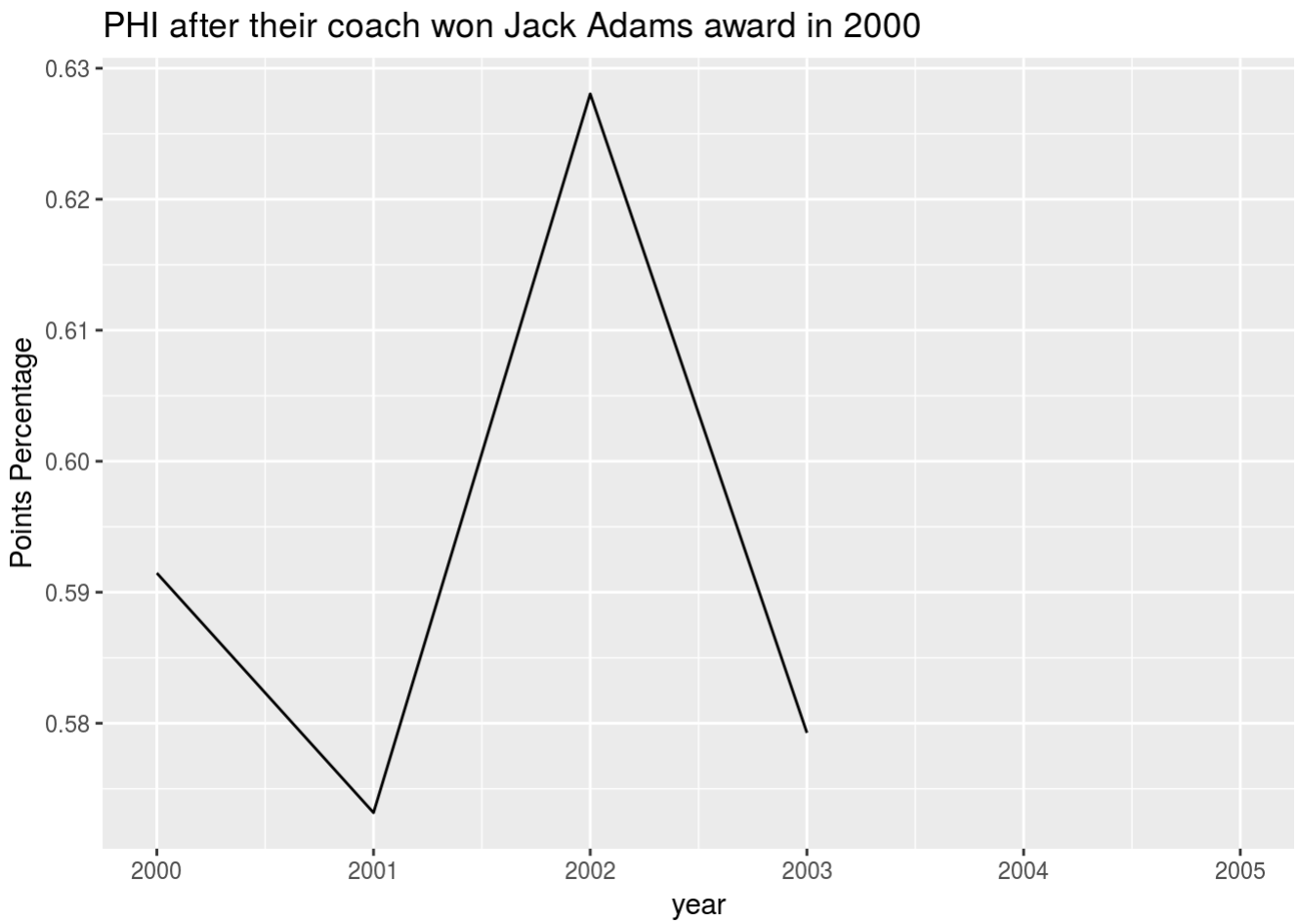
BUF after their coach won Jack Adams award in 1996

BOS after their coach won Jack Adams award in 1997


0.56
0.52
0.48
0.44
Points Percentage
1997
1998
1999
2000
2001
2002
year


OTT after their coach won Jack Adams award in 1998
Points Percentage
0.65
0.60
0.55
1998
1999
2000
2001
2002
2003
year


STL after their coach won Jack Adams award in 1999
Points Percentage
0.65
0.60
0.55
1999
2000
2001
2002
2003
year



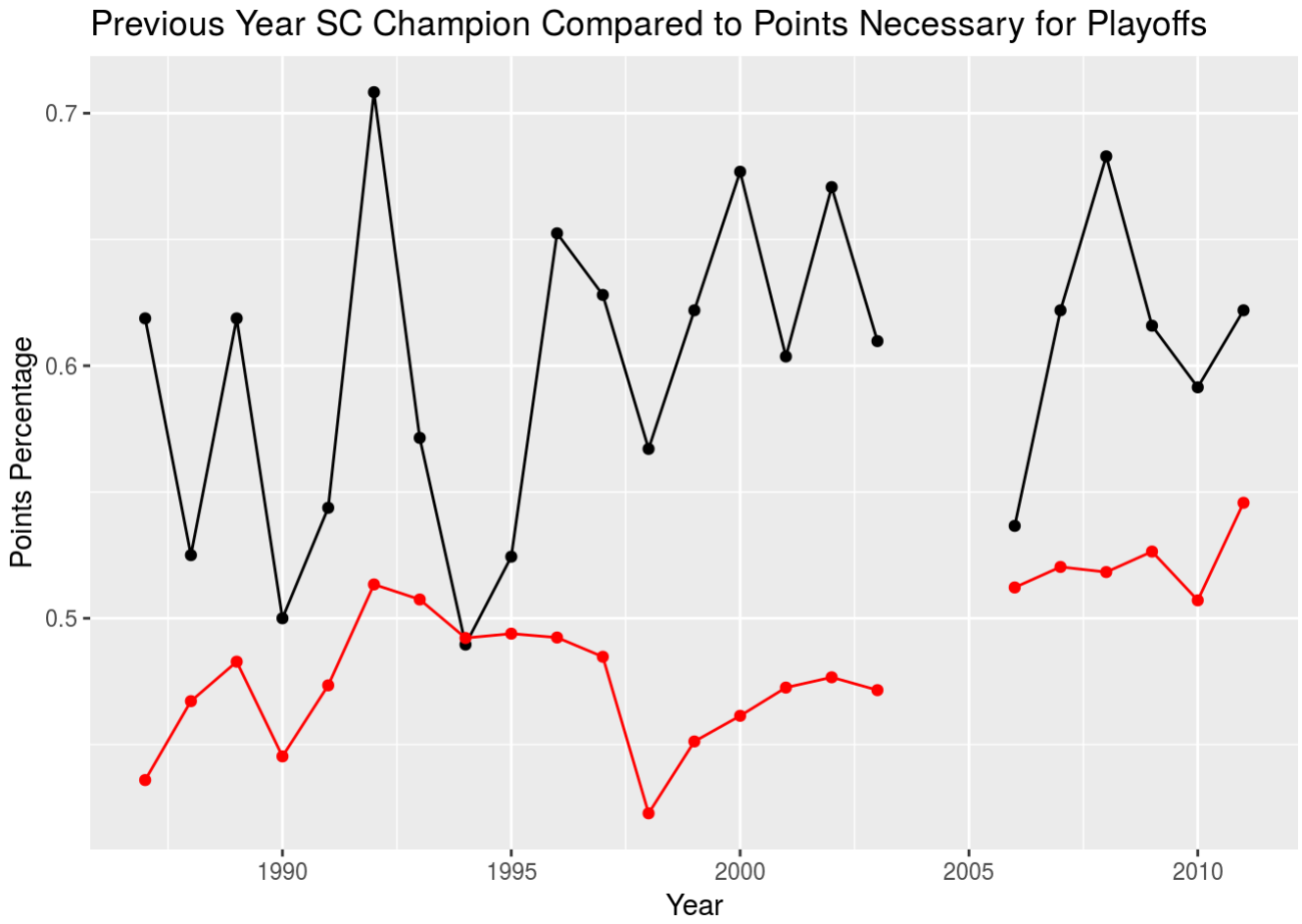
Question 2

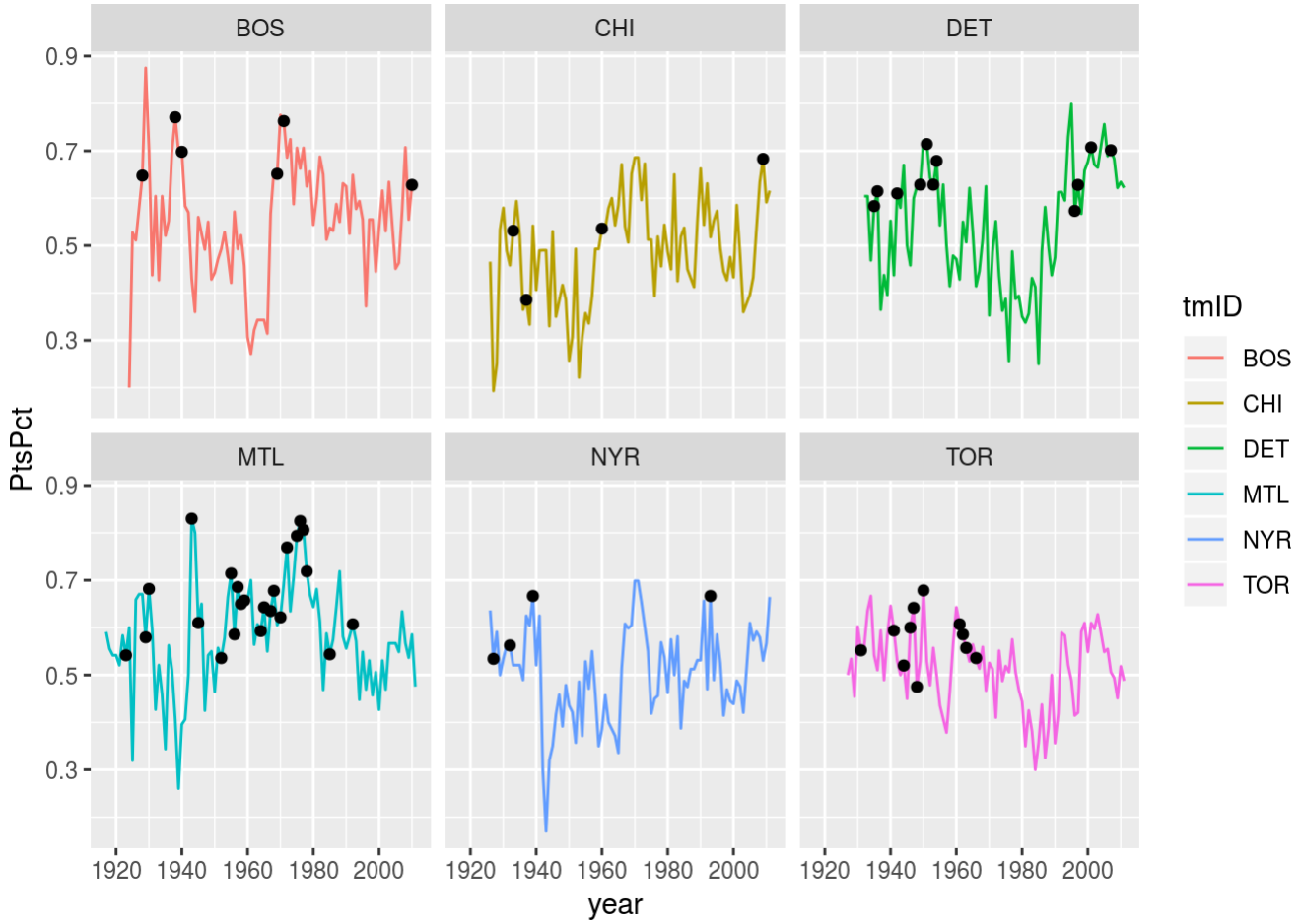
How does the team performance change after winning a Stanley Cup?

answer

In the graph below, the black line represents the percentage of total possible points per game that the team that won the Stanley Cup the previous year earns. The red line represents the minimum required number of points a team needs to make the playoffs. Therefore, since the average points per game for all but one year since 1987 is

higher for the reigning league champions than the rest of the NHL, team performance tends to remain high for teams that have just won a Stanley Cup the year before.





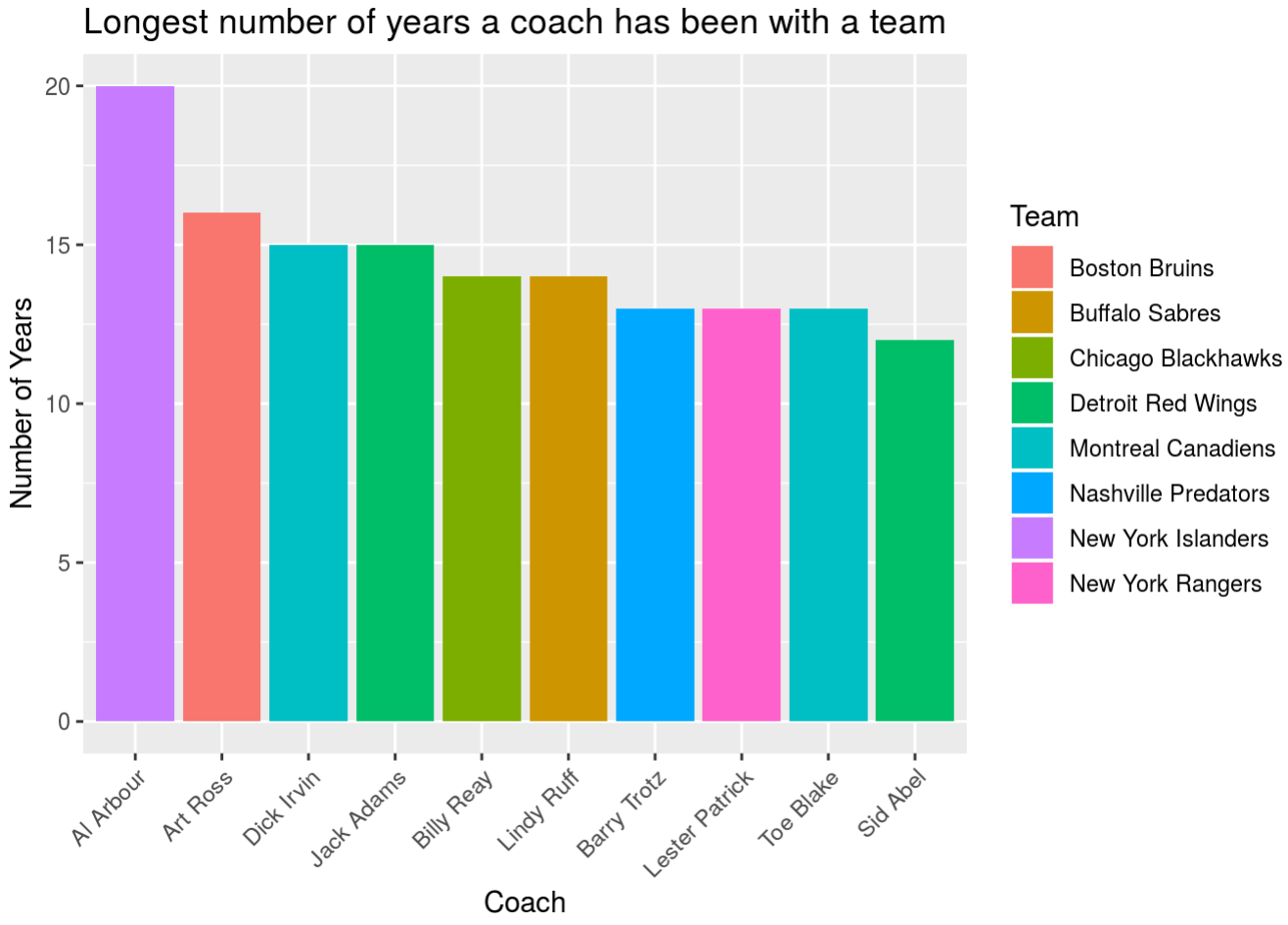
Question 3

What coach coached the same team for the longest number of years?

answer

Al Arbour has the record for coaching the same NHL team for the longest number of years (New York Islanders for 20 years), followed by Art Ross with the Boston Bruins for 16 years and Dick Irvins and Jack Adams (Montreal

Canadiens and Detroit Red Wings respectively) with 15 years each.



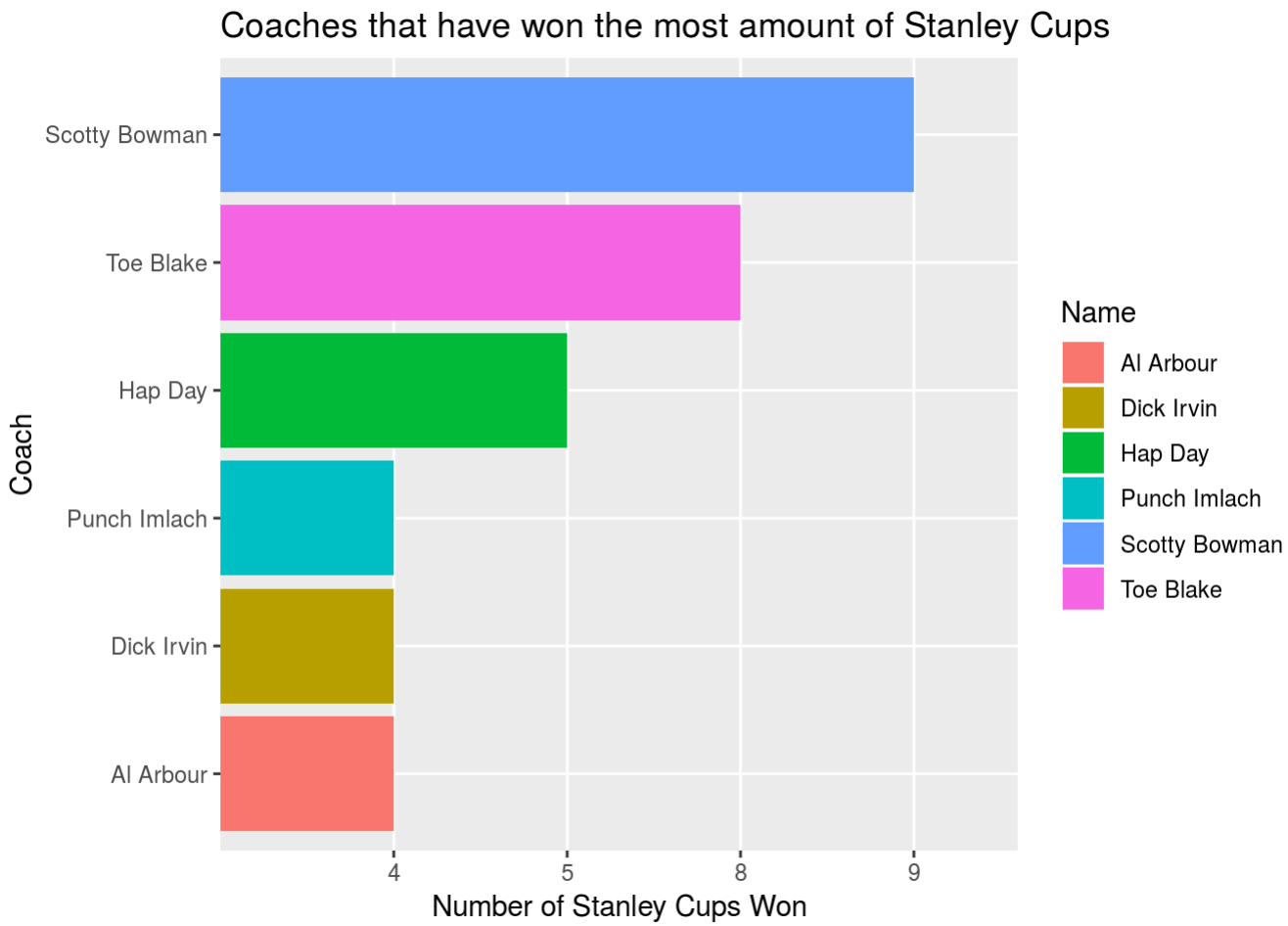
Question 4

What coach won the most Stanley Cups?

answer

Scotty Bowman has won the most Stanley Cups, followed by Toe Blake and Hap Day.

```
## # A tibble: 3 x 3
##   coachID        n Name
##   <chr>      <int> <chr>
## 1 bowmasc01c     9 Scotty Bowman
## 2 blaketo01c     8 Toe Blake
## 3 dayha01c       5 Hap Day
```



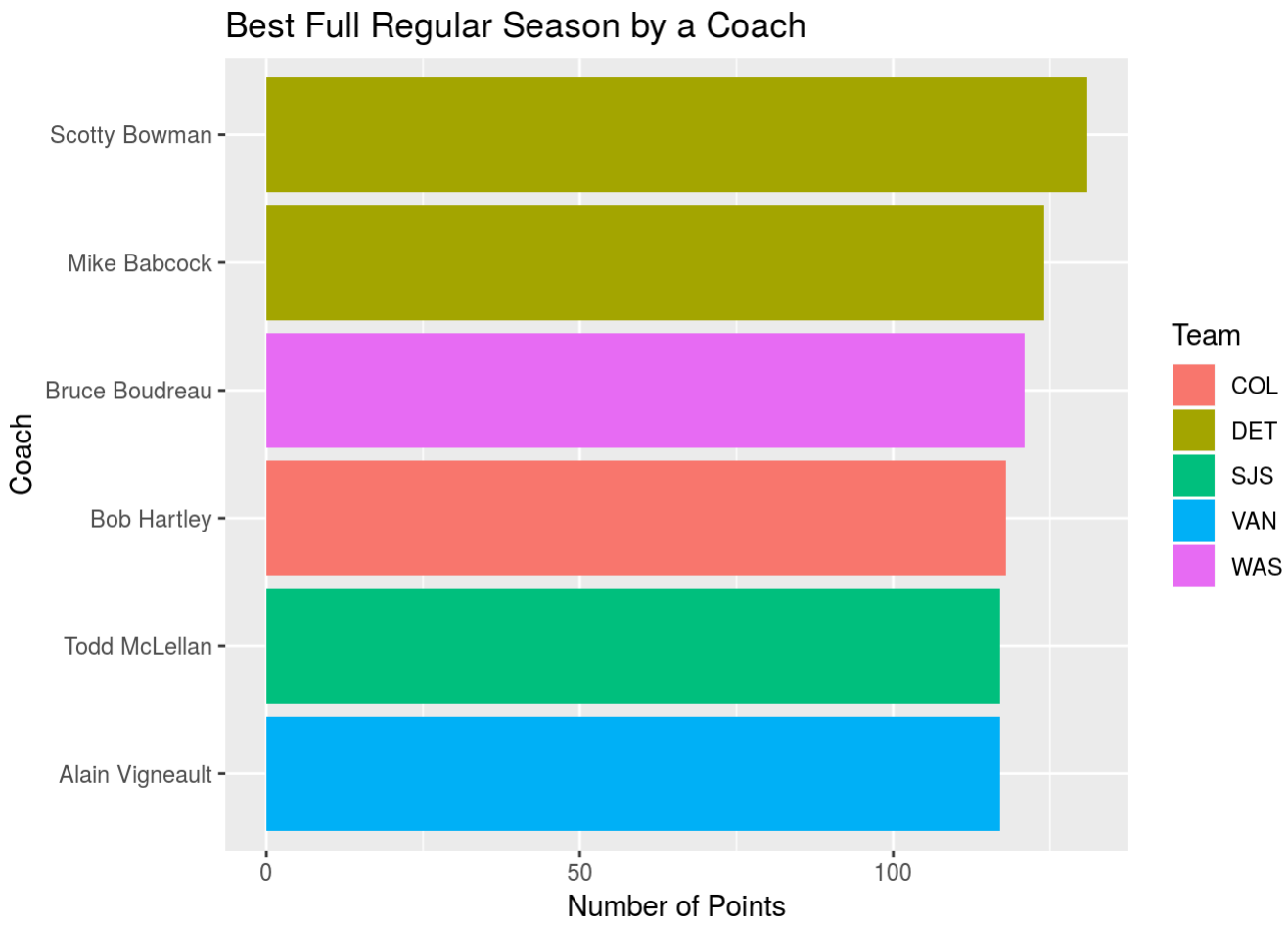
Question 5

What coach had the best full regular season (in terms of number of points per season)?

answer

The total number of possible points for a full regular season is 164 points (2 for a win, 1 for overtime game and 0 for a loss). Therefore, Scotty Bowman had the best full regular season with 131 points in 1995 with the Detroit Red Wings. He is followed by Mike Babcock in 2005 with the Red wings, and by Bruce Boudreau in 2009 with the Washington Capitals.

```
## # A tibble: 3 x 4
##   Name           tmID   year   Pts
##   <chr>          <chr> <dbl> <dbl>
## 1 Scotty Bowman  DET    1995   131
## 2 Mike Babcock   DET    2005   124
## 3 Bruce Boudreau WAS    2009   121
```

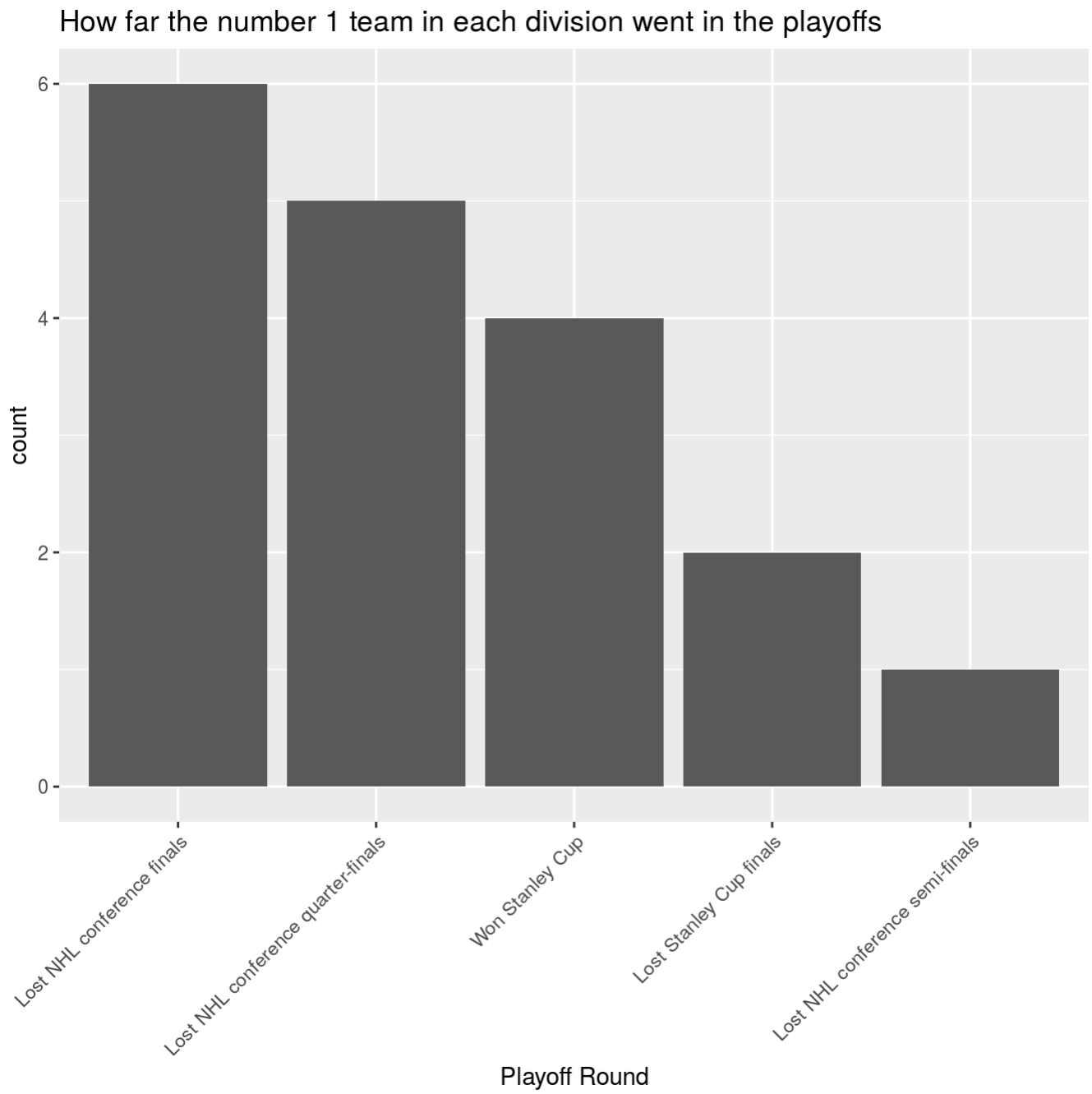


Question 6

Is there any correlation between regular season and post-season performance?

answer

From the data, we can see that the best team in the regular season isn't always the best team in the postseason. The majority of the time the best team doesn't even make it to the Stanley Cup finals.



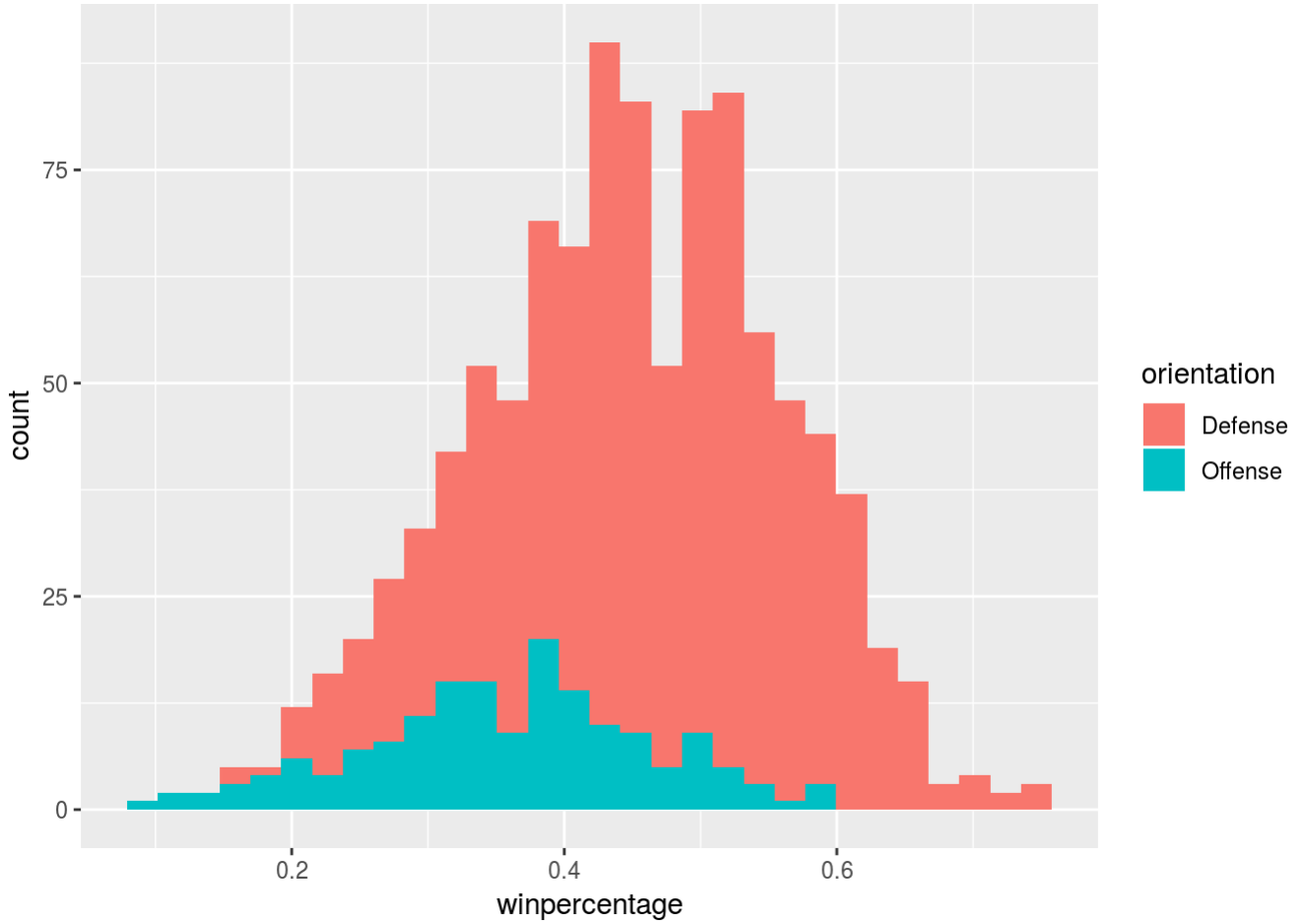
Question 7

Are there any coaching tactics that consistently result in a good record? High PKC / GA (Defensively oriented) or High PPG / GF (Offensively oriented)?

answer

It appears that there are much more defensively oriented teams than offensively oriented team. This is likely to be due to fact that defensively oriented teams throughout history have had higher mean and median win percentages than offensively oriented team.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Conclusion

Q1.

We tried to adopt a scalable approach when dealing with question 1. Rather than creating multiple tibbles, merging the data and using a faceted plot, we decided to use a function and loop. The loop iterates through a tibble of all the Jack Adam winners in a certain range. It then finds the respective team data from our main table and plots the performance of the team for the next 5 years. This approach allows us to quickly swap out the dataframe passed as an argument so we can tweak the range or other parameters without too much hassle. On the downside the plots take up individual outputs and are pretty bland.

Question 1

We tried to adopt a scalable approach when dealing with question 1. Rather than creating multiple tibles, merging the data and using a faceted plot, we decided to use a function and loop. The loop iterates through a tibble of all the Jack Adam winners in a certain range. It then finds the respective team data from our main table and plots the performance of the team for the next 5 years. This approach allows us to quickly swap out the dataframe passed as an argument so we can tweak the range or other parameters without too much hassle. On the downside the plots take up individual outputs and are pretty bland.

As touched on above we see that almost every team does worse the year after their coach wins a Jack Adams award. We have attributed this to the cumulative effects of team reorganization, expectations, regression effect and that fact that the probability that a team performs significantly well another year in a row is lower than performing averagely.

Question 2

For this question my first goal was to get a list of Stanley Cup champions for the modern playoff structure (16 wins in the postseason = Stanly Cup). To do this I filtered our post seaon data set for teams that got 16 wins, selected the relevant variables (year and tmID), then iterated the year by one so when we merged the dataset we will get information for year after they won the Stanley Cup. The basic metric I was planning on using is Points %. This is simply the number of points the team got, divided by the possible number of points the team could have got. After doing this using the regular season data, I joined the data sets by year and tmID to ensure I did not get any duplicate information. Having this, I wanted something to compare the data to. In order to do that I computed the points threshold to make it into the playoffs for each season. Having this I plotted each each data set and noticed that teams that had won the Stanley Cup the previous year made it back to the playoffs the next year in every season except one.

Question 3

Since coaches are free to move between teams, we thought a good question to ask was how loyal are coaches to their respective teams. In other words we wanted to find the coaches with the highest consecutive years with a team. Our initial approach to this was to subtract the max year value from the minimum year value and ensure that it was equal to the number of rows in the filtered tibble. This would ensure that those years were actually consecutive and the coach was with a team throughout. However, trying to get r to do this without manually creating multiple tibbles and comparing the difference to nrow() was quite tedious. In the end we opted to just see how long a coach has been with a particular team, regardless of whether they moved and returned.

When fact checking the stats we got we see that Al Arbour coached the New York Islanders from 1973–1986, left for two years and then returned to coach them from 1988-1994. Which indeed adds up to our total of 20 years. Art Ross coached the Boston Bruins from 1924-1945 taking two 2 year gapes in between.

Our next idea was to take the average performance of the teams under these coaches and compare it with the coaches after for the same teams. This however led to issues as the data was not complete and the coaches were not in the same time frame, leading to the recurring issue of the league being changed between seasons.

Question 4

Often times, people debate who is considered as the greatest of all time (GOAT). Usually, this is in the context of individual athletes. However, seeing as this project was looking at coaching data from North American Hockey, we wanted to see which coach can be considered the GOAT.

Question 5

A common theme amongst sports fans is to argue which team is better, when a lot of the time these arguments come down to a personal preference in which statistics to use. We thought that a good way to combat this problem was to look at which teams (in the sense that the ’95 Islanders is a different team than the ’96 Islanders) have been the most sucessful throughout NHL history. Since we were interested in coaching statistics, we looked into this question from a coaching standpoint, crediting the teams success to the coach at the time (although we certainly recognize that like all team sports, success can never one player/one coach).

In 1995, Scotty Bowman, who at the time was coaching the Detroit Red Wings, had the best full regular season to date in NHL history. His team earned 131 points out of 164 possible (79.9%). The second best full regular season was in 2005, also by the Detroit Red Wings under coach Mike Babcock (124 points). Bruce Boudreau has the third best full regular season with 121 points with the 2009 Washington Capitals.

Question 6

For this question we needed a way to filter for the best team during the regular season. To do this we grouped by the variable year then filtered for years 1994+ and the team that had the max number of points each year. From there we joined with the abbreviations data set to ensure we label our bars correctly. The abbreviations data set

contains numerous different abbreviations covering all facets of hockey, so to ensure we are only joining the relevant ones we filtered the data set for only the “playoff” abbreviations. From there we plotted how far each team went in the playoffs.

Question 7

A good way to ensure the most efficient team possible is to look at which tactics produce the best results. In this question, we were interested in seeing which overall strategies (offensively/defensively oriented) yield the highest win percentages in the NHL.

The qualification for being labeled as a defensively orriented team was chosen to have a penalty kill chance to goals against ratio above 1.2 or a power play goal to goals for ratio below .228. These values were chosen by sorting the created dataset after creating these ratios and looking for the approximate middle value of these ratios. From this, we see that there have been much more defensively oriented teams throughout history and that they have historically done better than their offensively oriented counterparts. For defensively oriented teams, their mean and median win percentage are 0.461 and 0.463 respectively, with a standard deviation of 0.107. For offensively oriented teams, their mean and median win percentage are 0.357 and 0.363 respectively, with a standard deviation of 0.105.

Final Thoughts

In respect to our methods of tidying and analysis: We ended up splitting our many datasets into even smaller dataframes with the desired variables. This left us with a working environment a little too full of dataframes. In the future it might be wiser to overwrite and only keep our desired dataframes rather than keeping both. It is also hard to find trends in this kind of sports data since the players that make up each team change every year. So the performance of a team varies much more than expected from season to season. In the future it might be a good idea to track the stats of individual players rather than just the entire performance of a team. This will result in more accurate and meaningful visualizations and interpretations. Another thing we could do in the future is pick a subset of years in which the league and its rules do not vary. This will allow us to make more objective conclusions and comparisons between coaches, teams and their respective performance.

-Our presentation consists of animations and overlaid images that do not appear correctly in the included PDF document. For the best viewing experience please visit https://docs.Google.com/presentation/d/1KHiInxtCBDgtVvIivzRppeJThCrnS44XxOqWlmDvWJA/edit?usp=sharing (https://docs.Google.com/presentation/d/1KHiInxtCBDgtVvIivzRppeJThCrnS44XxOqWlmDvWJA/edit?usp=sharing)