An Exploration of NHL Coaching Statistics

Tanush Samson, Zach Morales & Pete Schultz

Data Sources

www.kaggle.com/open-source-sports/professional-hockey-database

Data & Scope

Data contained information about the NHA, PCHA, NHL, WCHL and the WHA.

Ranges from 1917 to 2011.

Uses a custom set of keywords & abbreviations for entries such as awards.
(SC for won Stanley Cup, F for lost Stanley Cup finals, etc)

We decided to work with NHL data with a normalized point percentage to account for the variation in games played per season over the years.

Research Questions

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

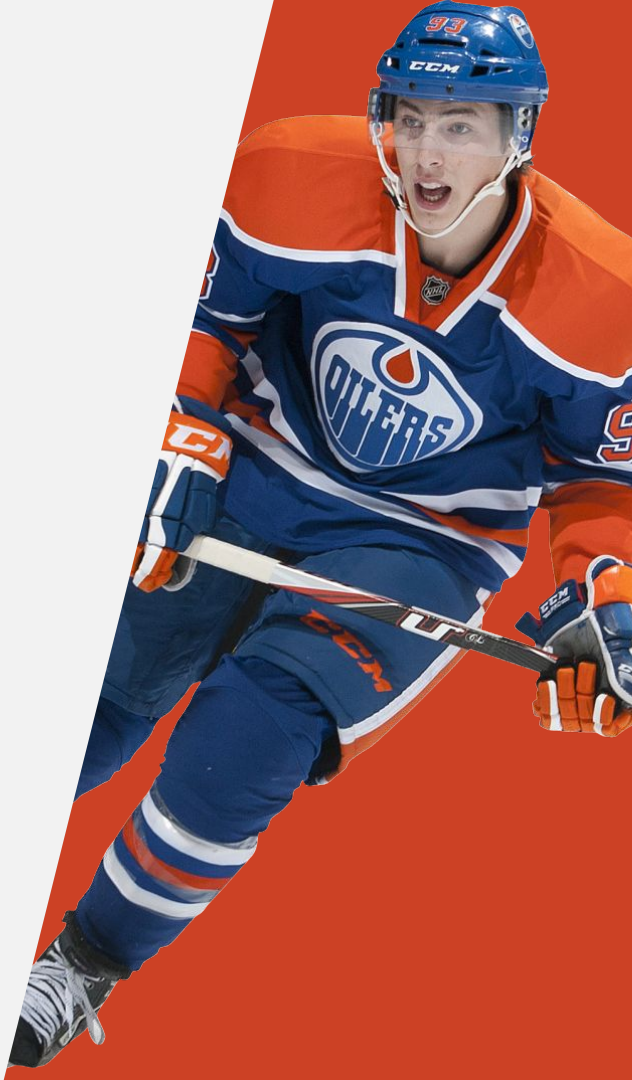
How does team performance change after winning a Stanley Cup?
(Stanley Cup hangover? Salary cap?)

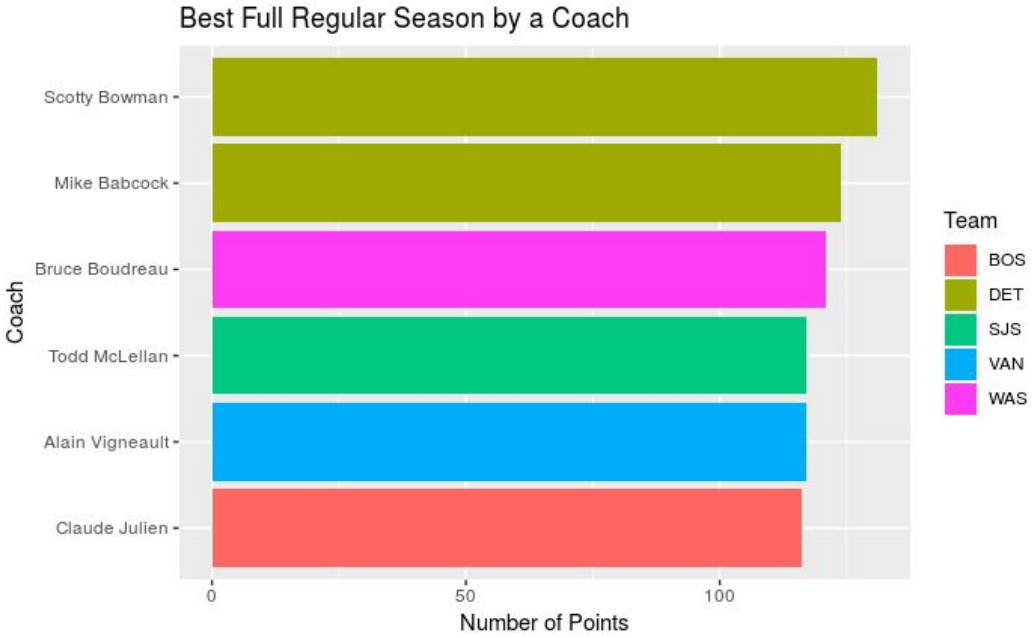
What coach coached the same team for the longest number of years?
(How does team performance vary after the coach gets an award?)

Question 1

What coach had the best full regular season?

```r
question5 = coaches %>%
  full_join(teams, by = c("year","tmID","lgID")) %>%
  filter(lgID == "NHL",g == 82) %>%
  select(coachID,year,tmID,lgID,g,w,t) %>%
  mutate(num_of_points = 2*w +t) %>%
  left_join(masters, by = "coachID") %>%
  select(firstName,lastName, tmID,year,num_of_points) %>%
  unite(Name,firstName,lastName, sep = " " ) %>%
  arrange(desc(num_of_points))
```





Name <chr>	tmID <chr>	year <dbl>	num_of_points <dbl>
Scotty Bowman	DET	1995	131
Mike Babcock	DET	2005	124
Bruce Boudreau	WAS	2009	121

3 rows

Question 1 Summary

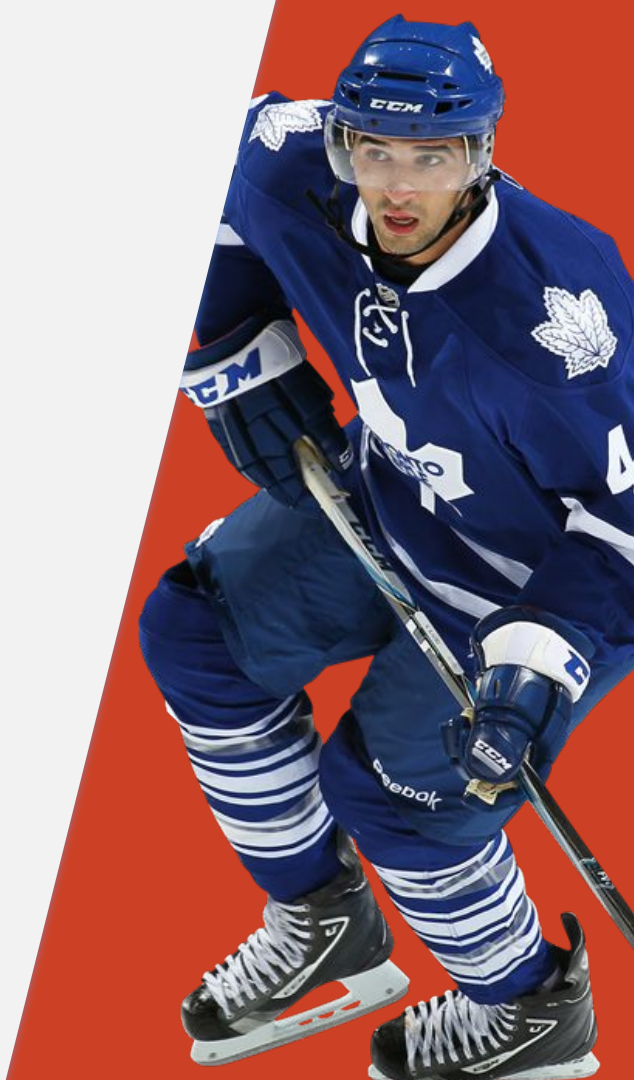
Scotty Bowman had the best regular season with 131 points in 1995.

Mike Babcock had the second highest with 124 points in 2005.

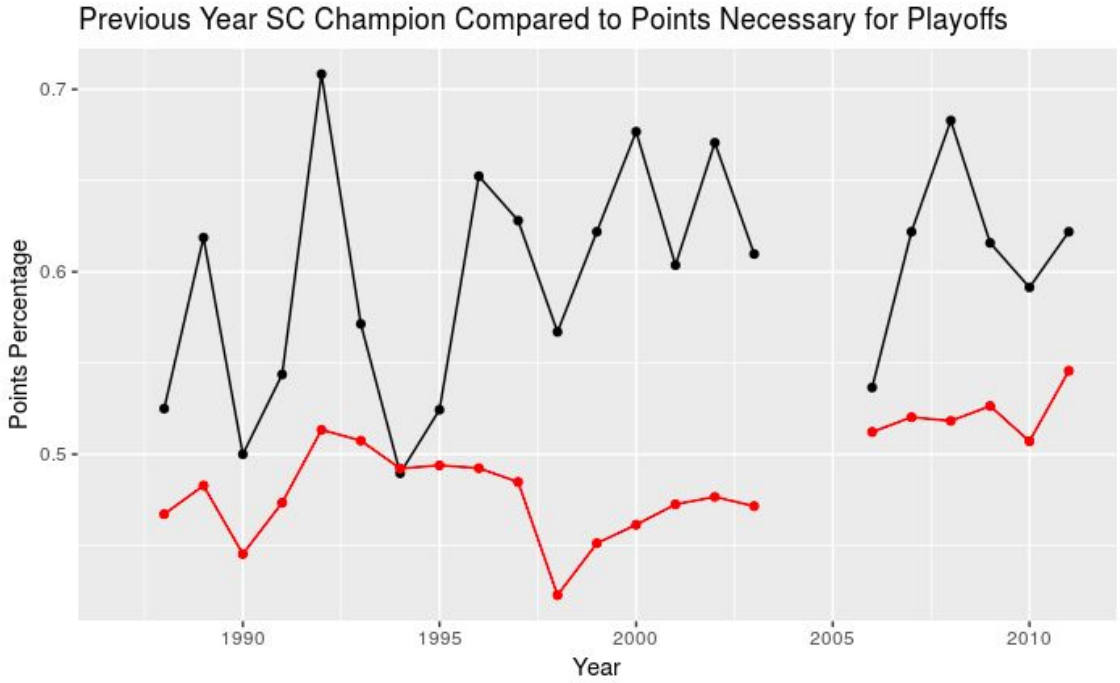
Bruce Boudreau came in third with 121 points in 2009.

Question 2

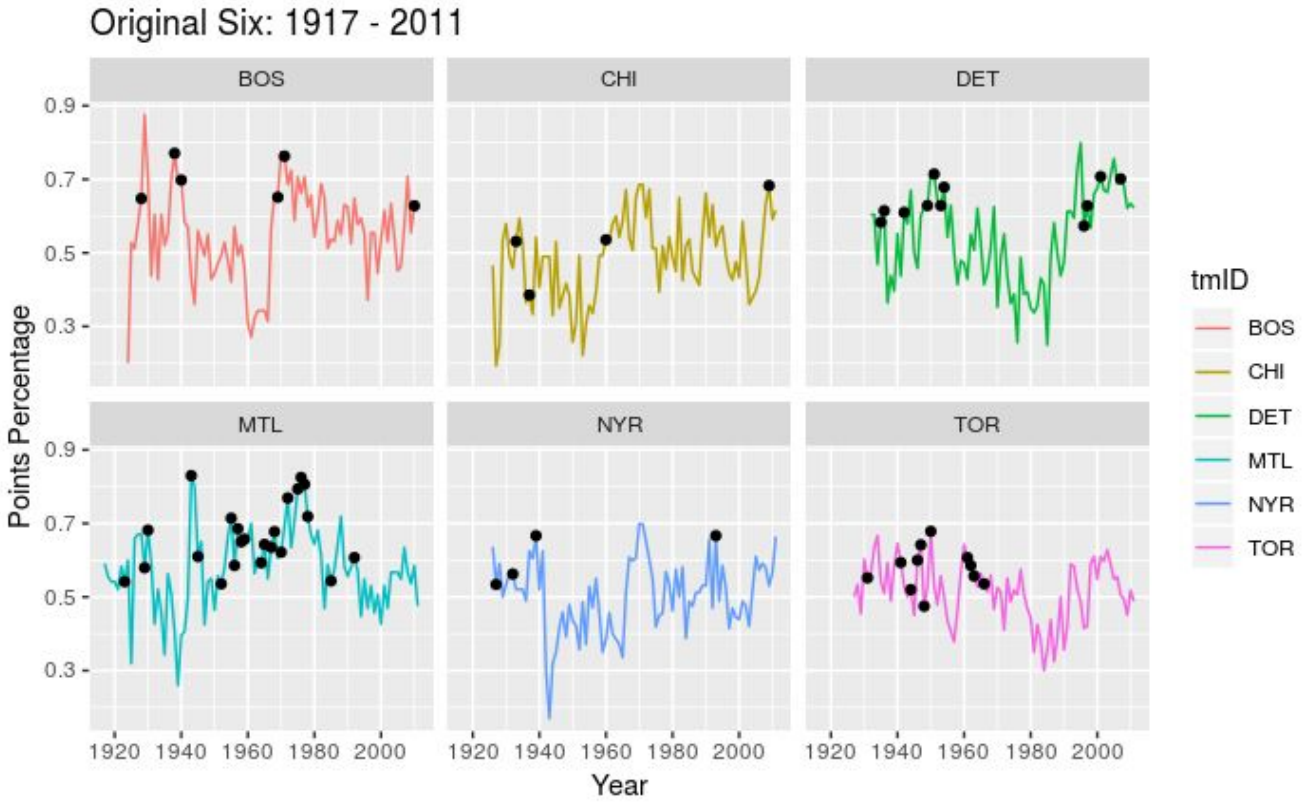
```
#### Question 2

#stantley cup champion, year + 1 to be able to join with another dataset
#to see how they do the next year

#filtering for SC champs in modern playoff structure (needs 16 wins to get Stanley cup)
scChamps <- teamsPost %>%
  filter(year < 2011, W == 16) %>%
  select(year, tmID) %>%
  mutate(year = year + 1)

#filtering reg season data and adding a column for Pts %.
#Pts % is the num of pts a team got during the season divided by the total points possible

regSeason <- teams %>%
  select(year, tmID, Pts, G) %>%
  filter(year > 1987) %>%
  group_by(year) %>%
  mutate(PtsPct = Pts/(2*G))

#merging the data sets
data_f <- left_join(scChamps, regSeason, by = c("year","tmID"))

#points needed for playoffs removing 2004 & 2005 as lockout years
playoffpts <- teams %>%
select(year, tmID, Pts, G, rank) %>%
filter(year > 1987, rank == 4, year != 2004 , year != 2005) %>%
group_by(year) %>%
summarise(PtsForPlayoffs = mean(Pts/ (2*G)))

#joining the data sets to plot our data
data_f <- left_join(data_f, playoffpts)
```



```
ggplot(data_f) +
  geom_line(aes(year, PtsPct), col = "black") +
  geom_point(aes(year, PtsPct), col = "black") +
  geom_line(aes(year, PtsForPlayoffs), col = "red") +
  geom_point(aes(year, PtsForPlayoffs), col = "red") +
  labs(title = "Previous Year SC Champion Compared to Points Necessary for Playoffs",
       x = "Year", y = "Points Percentage")
```



```{r}
#filtering for orignal six, selecting relovant variable, creating pts %
origSix <- teams %>%
  filter(tmID == "BOS" | tmID == "MTL" | tmID == "DET" |
           tmID == "CHI" | tmID == "TOR" | tmID == "NYR") %>%
  select(year, tmID, G, Pts, playoff, name) %>%
  mutate(PtsPct = Pts / (2*G))

#filtering for the years each team won the stanley cup
SC <- origSix %>%
  filter(playoff == "SC")

#plotting the data
ggplot(origSix, aes(year, PtsPct, col = tmID)) +
  geom_line() +
  geom_point(data = SC, aes(year, PtsPct), col = "black") +
  facet_wrap(~tmID)
```


Original Six: 1917 - 2011
BOS
CHI
DET
MTL
NYR
TOR
Points Percentage
0.9
0.7
0.5
0.3
1920
1940
1960
1980
2000
Year
tmID
BOS
CHI
DET
MTL
NYR
TOR


91
Reebok

Question 2 Summary

After a season where a team wins the Stanley Cup, they are able to provide another successful season.

There is only one case where a team failed to make the playoffs after winning a Stanley Cup.

Question 3

What coach coached the same team for the lo

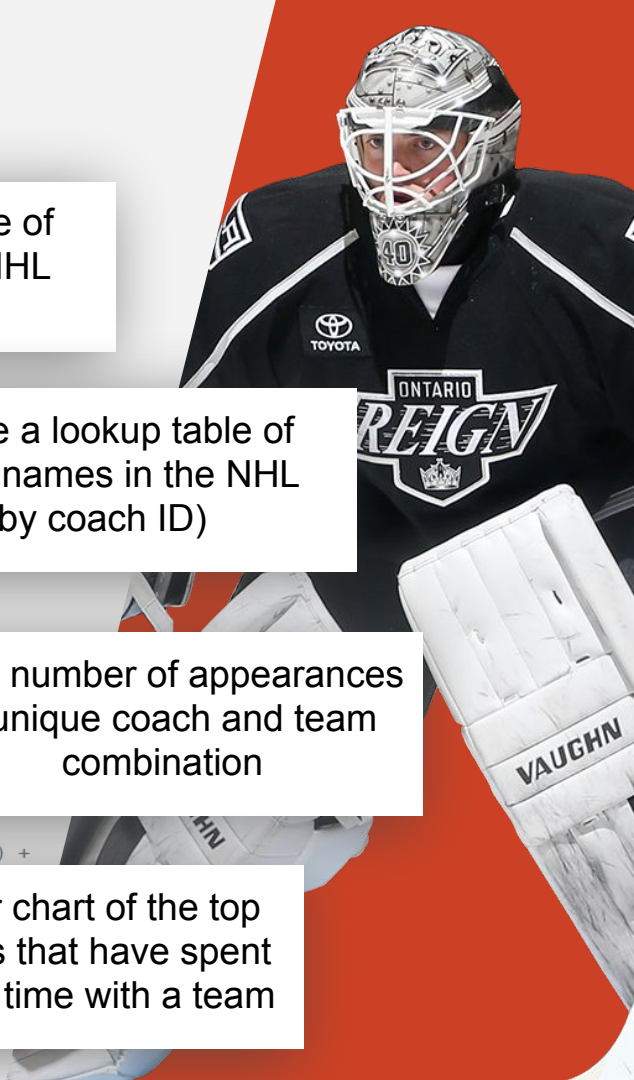
```r
#Cleaning up a team name lookup table
teamNames <- teams_F %>% select(c("tmID", "name")) %>%
             unique() %>% filter(name!="Chicago Black Hawks")

#Cleaning up a coach name lookup table
coaches_names <- masters %>% select("coachID", "firstName", "lastName") %>%
                 filter(!is.na(coachID))

#Join coach data with coach names
topCoaches<- coaches %>% group_by(coachID, tmID) %>%  tally() %>% arrange(desc(n)) %>%
             inner_join(., coaches_names, by="coachID") %>%
             mutate(nameC = paste(firstName, lastName, sep=" ")) %>% head(10)

#Joing team name with tibble of top coaches
topCoaches %>% inner_join(., teamNames, by="tmID") %>%
  ggplot(aes( reorder(nameC, -n),n)) + geom_col(aes(fill=name)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Longest number of years a coach has been with a team", y="Number of Years", x="Coach") +
  scale_fill_discrete(name="Team")
```

Create a lookup table of team names in the NHL (by team ID)

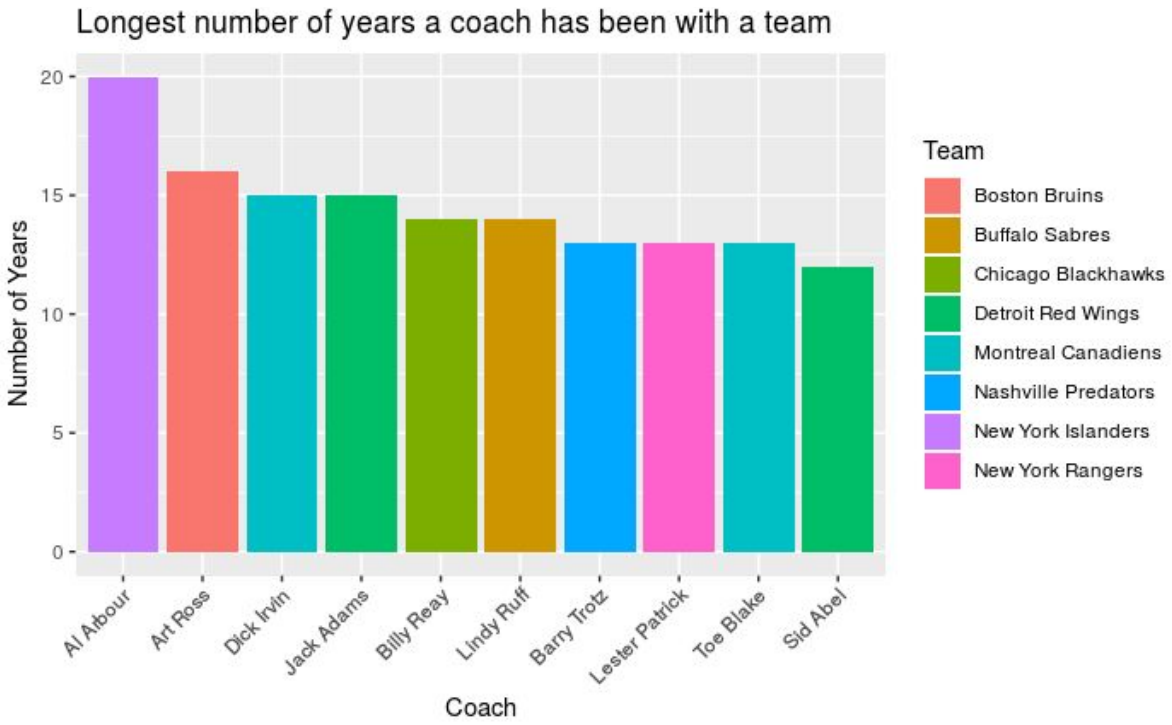
Create a lookup table of coach names in the NHL (by coach ID)

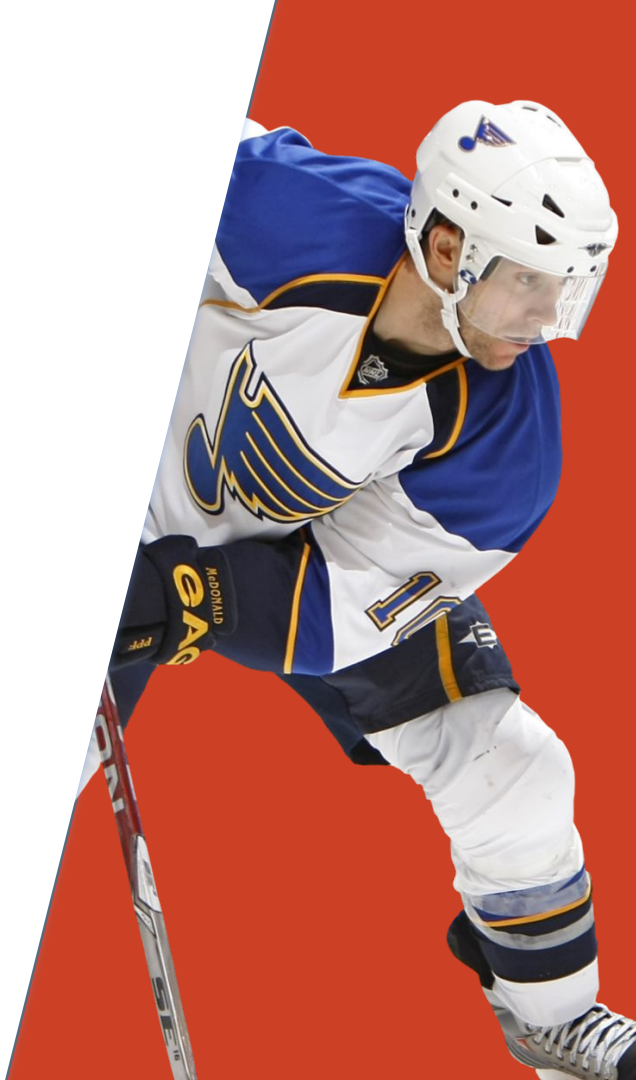
Tally number of appearances of unique coach and team combination

Create bar chart of the top 10 coaches that have spent the longest time with a team



Question 3


Longest number of years a coach has been with a team
Number of Years
20
15
10
5
0
Al Arbour
Art Ross
Dick Irvin
Jack Adams
Billy Reay
Lindy Ruff
Barry Trotz
Lester Patrick
Toe Blake
Sid Abel
Coach
Team
Boston Bruins
Buffalo Sabres
Chicago Blackhawks
Detroit Red Wings
Montreal Canadiens
Nashville Predators
New York Islanders
New York Rangers



Question 3

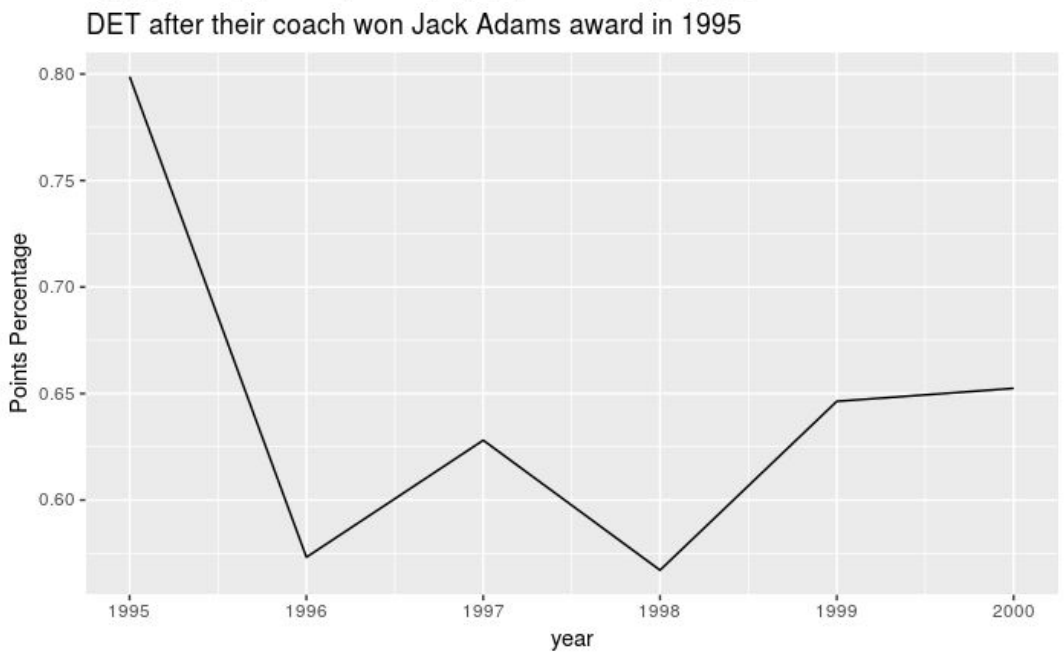
How does team performance vary after the coach gets an award?

The Jack Adams Award is awarded annually to an NHL coach who has been determined to have contributed the most to his team's success.

Started in 1974, the league's Coach of the Year award has been presented 40 times to 34 coaches.

Question 3

How does team performance vary after the coach gets an award?



Question 3 Summary

Al Arbour coached the New York Islanders for 20 years.

Art Ross coached the Boston Bruins for 16 years.

Dick Irvin and Jack Adams coached the Montreal Canadians and the Detroit Red wings for 15 years respectively.

Teams seem to almost always do worse the season after a coach wins a Jack Adams award.
(go figure!)

Problems faced

Data was scattered between 20 datasets.

Each dataset contained cases from 4 other leagues as well which resulted in a large number of variables per table as well as a lot of NA values.

Since the rules around the league have changed since our data begins in 1917, there are many cases where values are not consistent.
(Such as number of games played in a season)

Questions?