

HW 8: NHL stats

Stats and sports class

Fall 2019

```
library(RCurl); library(tidyverse)
gitURL<- "https://raw.githubusercontent.com/statsbylopez/StatsSports/master/Data/pbp_data_hockey.rds"
pbp_data <- readRDS(gzcon(url(gitURL)))
```

Question 2

McDavid is credited with 14.3 expected goals off of his passes. In reality, he finished the 20182019 season with 75 assists. How come McDavid has so few expected assists in this data set? Think carefully about how NHL data is collected. If you aren't sure, give a guess! But also explore the data to see if you pick anything up.

Assists only show up when players actually score the goal – on other passes that set up shots (which we would want to understand passing skill), no event_player_2 is shown

Question 3/4

A coach wants to know if players can consistently overperform the average finishing of an NHL players. Make a plot of each players' goals above expectation in the 2018 season (+ or -) versus his goals above expectation in the 2019 data set. As a reminder, below is code from class which gets you started. You'll need to make two new variables that correspond to goals above or below expectation in each season, and compare them graphically.

Is there any positive link between the two new variables you created for performance above expectation? Use a smoothed trend curve and/or linear regression to make your conclusion.

```
season_2018 <- pbp_data %>%
  filter(season == 20172018) %>%
  group_by(event_player_1, season) %>%
  summarise(n_goals_18 = sum(event_type == "GOAL"),
            n_xGs_18 = sum(shot_prob),
            n_shots_18 = n()) %>%
  filter(n_shots_18 >= 100) %>%
  select(-season)

season_2019 <- pbp_data %>%
  filter(season == 20182019) %>%
  group_by(event_player_1, season) %>%
  summarise(n_goals_19 = sum(event_type == "GOAL"),
            n_xGs_19 = sum(shot_prob),
            n_shots_19 = n()) %>%
  filter(n_shots_19 >= 100) %>%
  select(-season)

season_combine <- season_2018 %>% inner_join(season_2019)

season_combine <- season_combine %>%
  mutate(gpm_18 = n_goals_18 - n_xGs_18,
         gpm_19 = n_goals_19 - n_xGs_19)
```

```
ggplot(season_combine, aes(x = gpm_18, y = gpm_19)) +
  geom_point() +
  geom_smooth()

fit_1 <- lm(gpm_19 ~ gpm_18, data = season_combine)
library(broom)
tidy(fit_1)
```

Solutions: In the scatter plot, there's a small, positive link between performance above average in 18 and performance above average in 19. Specifically, each goal above expectation in 2018 is linked to 0.27 goals above expectation in 2019. Note – either the regression or the scatter plot is sufficient

Question 6

Identify the one player who was 10 goals better than expectation in 2018 and 10 goals better than expectation in 2019.

```
season_combine %>% filter(gpm_18 >= 10, gpm_19 >= 10)
```

Solutions: Alex Ovechkin