

Econ 1042 PS2

Owen Asnis

2023-02-13

Data Wrangling

```
players <- read_delim("nhlps2.csv", delim = "\t",  
  escape_double = FALSE, trim_ws = TRUE)  
  
standings <- read_csv("standings_2018_2019.csv",  
  show_col_types = FALSE) %>%  
  select(-c(...11, ...12, ...13, ...14, ...15))
```

Question 1

In baseball, expected home runs is a better representation than actual home runs because the expected statistics account for other variables. For example, one batter could be making hard contact but only have a few home runs because they play in Pittsburgh (ranked by Bleacher Report as the worst ballpark for hitters), whereas another batter could be making softer contact but have more home runs because they play in Colorado (considered to be the best ballpark for hitters). Even though the player from Pittsburgh has less home runs, we would still consider them to be a better hitter, as they would have more home runs than the player from Colorado if they played in the same conditions. In hockey, the total shots by each team while a player is on the ice might be a better indicator of a player's value than goals and assists because it similarly accounts for other variables. For example, one player could be generating many shots for and allowing few shots against but have scored no goals because their team was playing the best goalie in the league, whereas another player could be generating few shots for and allowing many shots against and scored a goal because their team was playing the worst goalie in the league. Again, the player generating many shots for and few shots against would be considered the better player because they would be more productive than the player generating few shots for and many shots against if they were playing in the same conditions. In basketball, thinking about shots doesn't work in the same way, because after one team scores, the other team gets the ball, leading to very similar shot totals. In hockey, if one team is dominating the other, there will be a large difference in shot totals.

Question 2

Corsi is the net difference in shots taken by your team and the other team, as defined by the problem set. Therefore, this variable is trying to measure how good a player is at generating shots for their team and limiting shots for the team they're playing. Like was said in the response to Question 1, Corsi is useful compared to just goals scored at both the team and the player level, because it accounts for other variables, like talent of the goalie, and shows how good a player is at producing offense and providing defense beyond just simple goal statistics.