Statistical Analysis Report

Molly McNamara

Initial Evaluation of the Dataset

The cleaned dataset of NHL players and teams from the data wrangling report was used for this exploration. The R package "countrycode" was used to generate region (by continent) for player origins from the Birth_Country variable. A sub-dataset was generated that used only the first instance of each unique player in order to evaluate their demographics without multiple years of play affecting the statistics.

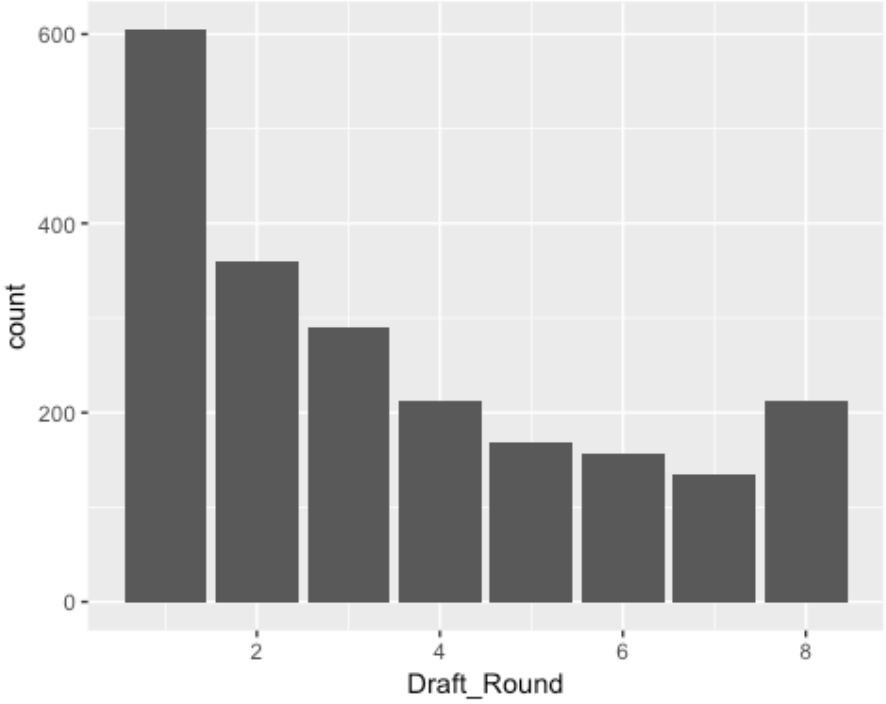
The full datset was initially evaluated with the summary command to review the variables and their basic statistics. See Appendix 1.

Basic statistics about players and their draft information were viewed next.

```
##      Height          Weight      FirstNHLseason LastNHLseason
##  Min.   :66.00   Min.   :144.0   Min.   :1979   Min.   :2000
##  1st Qu.:72.00   1st Qu.:190.0   1st Qu.:1996   1st Qu.:2003
##  Median :73.00   Median :200.0   Median :2002   Median :2009
##  Mean   :73.21   Mean   :201.5   Mean   :2001   Mean   :2007
##  3rd Qu.:74.25   3rd Qu.:211.0   3rd Qu.:2007   3rd Qu.:2011
##  Max.   :81.00   Max.   :265.0   Max.   :2011   Max.   :2011
##    Draft_Pick       Draft_Year     Draft_Age      Draft_Round
##  Min.   :  1.00   Min.   :1979   Min.   :17.00   Min.   :1.000
##  1st Qu.: 27.00   1st Qu.:1994   1st Qu.:18.00   1st Qu.:1.000
##  Median : 70.00   Median :1999   Median :18.00   Median :3.000
##  Mean   : 91.08   Mean   :1998   Mean   :18.69   Mean   :3.492
##  3rd Qu.:144.00   3rd Qu.:2004   3rd Qu.:19.00   3rd Qu.:5.000
##  Max.   :291.00   Max.   :2010   Max.   :31.00   Max.   :8.000
##    Draft_Pick
##  Min.   :  1.00
##  1st Qu.: 27.00
##  Median : 70.00
##  Mean   : 91.08
##  3rd Qu.:144.00
##  Max.   :291.00
```

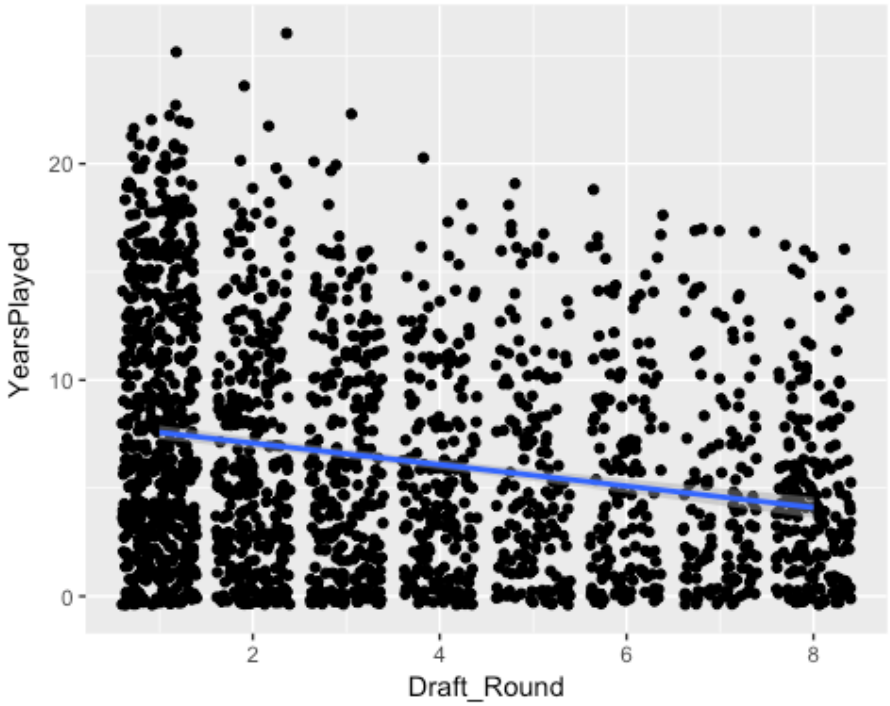
Draft Statistics

Draft statistics were initially evaluated by viewing a distribution of the draft rounds of the players in this dataset.



For active players between 2000-2012, the round they were drafted in appears to be skewed in one direction, towards players drafted in earlier rounds. It may be that players drafted in earlier rounds play longer in the league and therefore are more highly represented.

To further investigate that point, a comparison was made between draft round and active years (up to 2012) in the NHL for players in this dataset.



It does indeed appear that players from earlier draft rounds may have somewhat longer NHL careers.

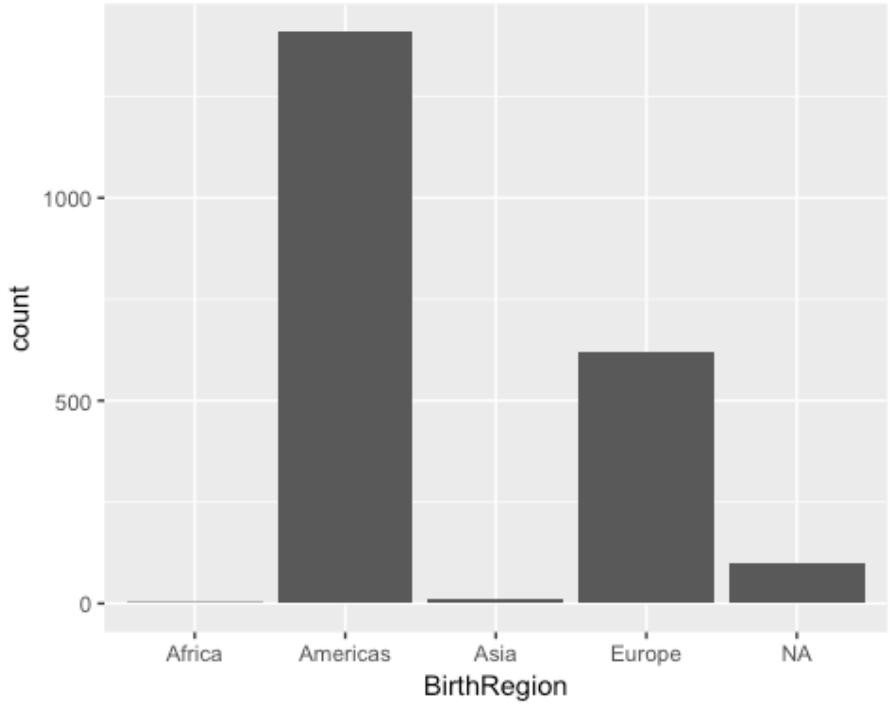
Draft and Country of Origin

To further explore factors around the NHL draft, tables were generated to summarize draft round and draft age by birth country. See Appendix 2.

Mean draft age looks to be similar across most countries, though this can be examined further later.

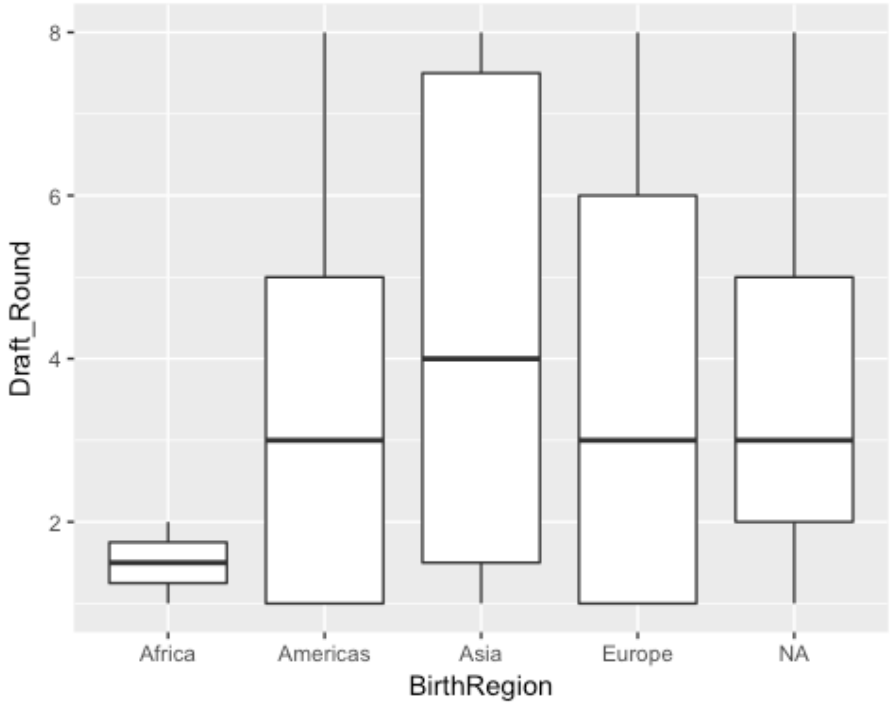
While it appears drafted players originate from many different countries, the NHL is based in the US and Canada; it might be useful to see regionally where the most players come from.

Warning: Ignoring unknown parameters: binwidth, bins, pad



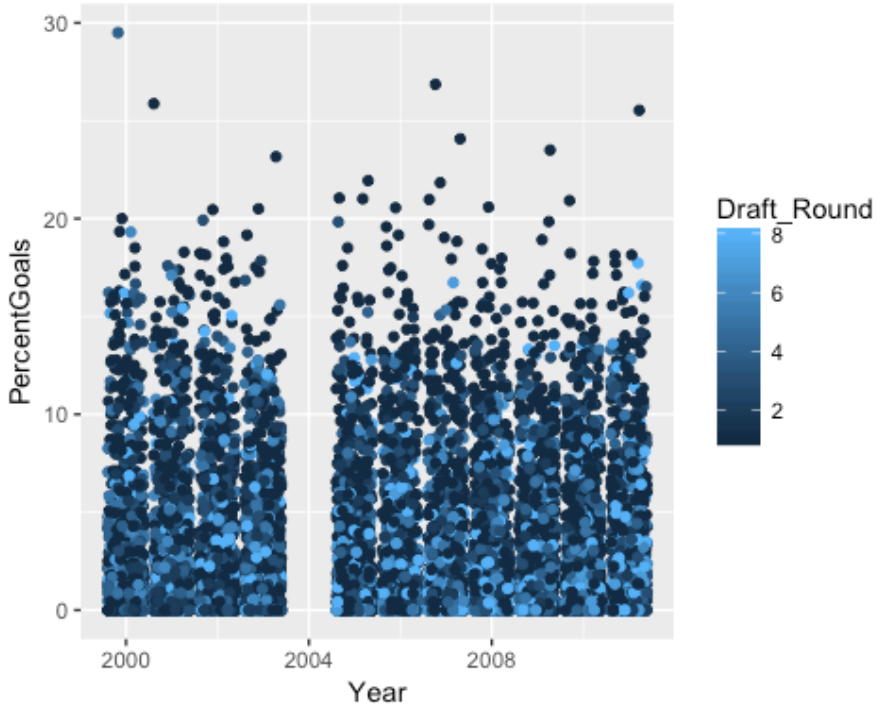
The Americas and Europe generate the vast majority of players drafted in the NHL.

Next it might be of interest to see if players from the various regions are drafted earlier or later than one another.



There may be some differences between regions. However, as seen in the prior graph, the majority of players come from the Americas and Europe, which appear to be distributed similarly here.

Players who are taken as high draft picks might be expected to make a greater contribution to their team. One way to look at this might be to evaluate if players who are drafted in earlier rounds score a greater percentage of their teams' goals.

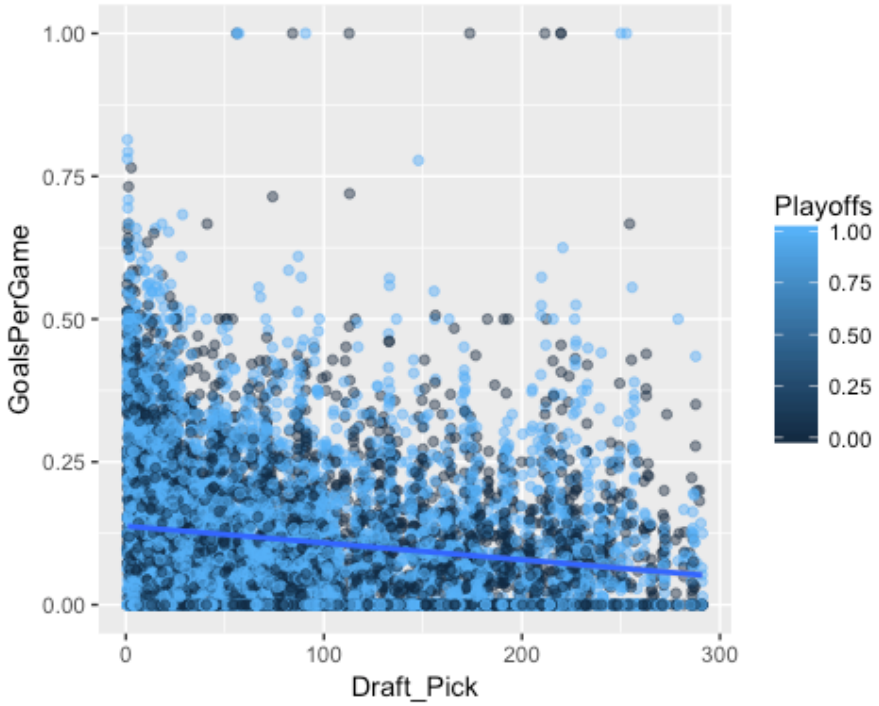


From this graph it does appear the players who score the highest percentage of their teams' goals were drafted in the earliest rounds but the relationship is less clear in the bulk of the data.

One might note that the data is missing for 2004 on this graph. Further investigation reveals that there was a leaguewide lockout in 2004 and no NHL games were played during a labor dispute between the owners and players.

Another way to evaluate the contribution of players drafted in earlier rounds might be to see if they score more goals per game and if this helps their team make the playoffs.

Warning: Removed 23 rows containing non-finite values (stat_smooth).
Warning: Removed 23 rows containing missing values (geom_point).

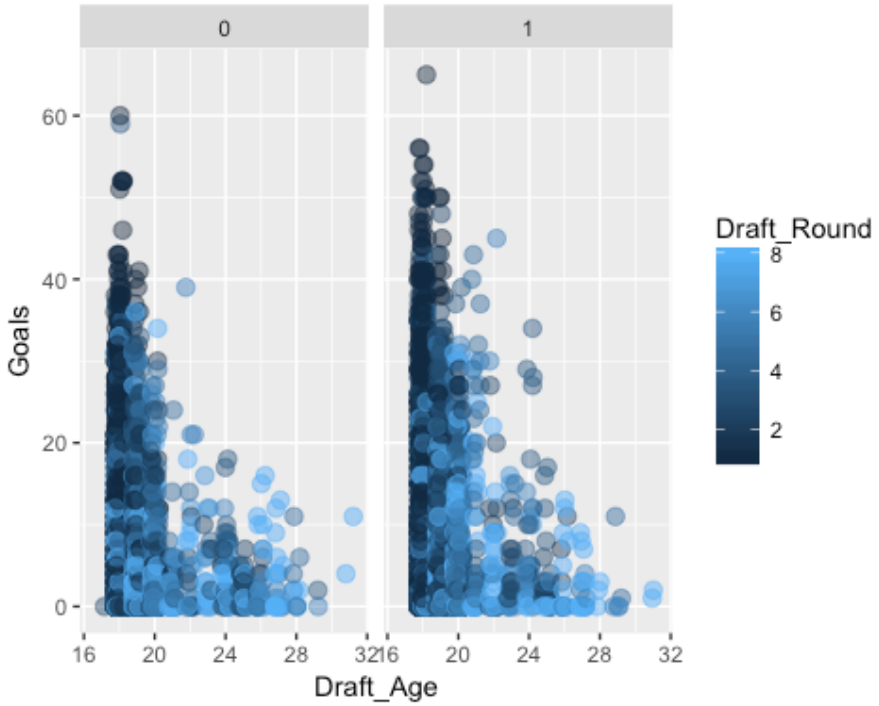


It does look like there is a slight trend for more goals per game from higher-drafted players but it's not clear if this impacts making the playoffs.

Player age at draft is another variable that might impact goal scoring and teams making the playoffs.

```
##    Draft_Age
##  Min.   :17.00
##  1st Qu.:18.00
##  Median :18.00
##  Mean   :18.69
##  3rd Qu.:19.00
##  Max.   :31.00
```

It appears that most players are drafted in a very narrow range, but exceptions do occur (31 years old in one case!). The next graph displays how draft age and round are related to goal scoring and teams making or not making the playoffs.

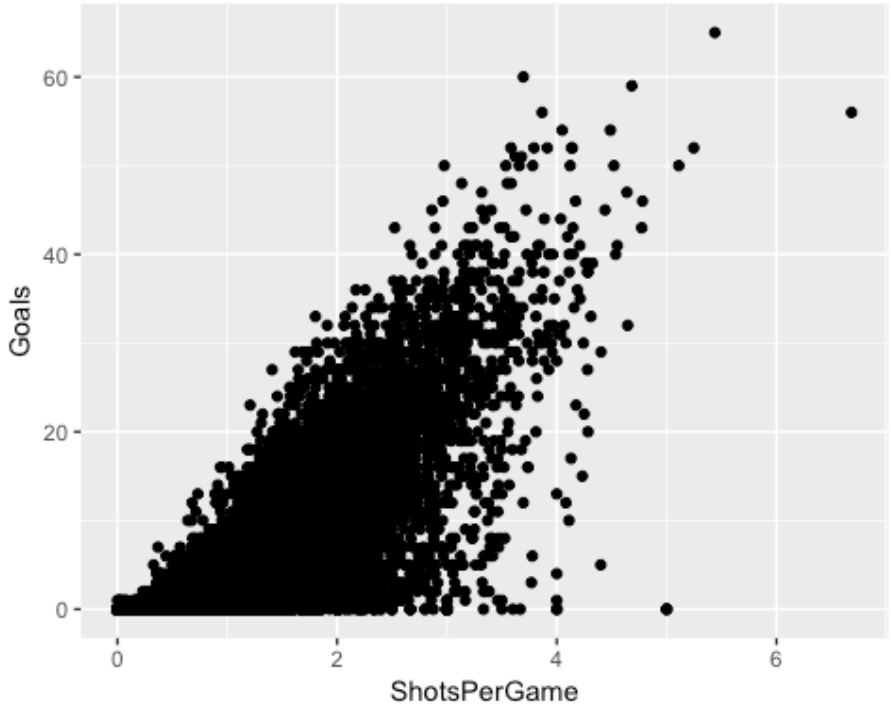


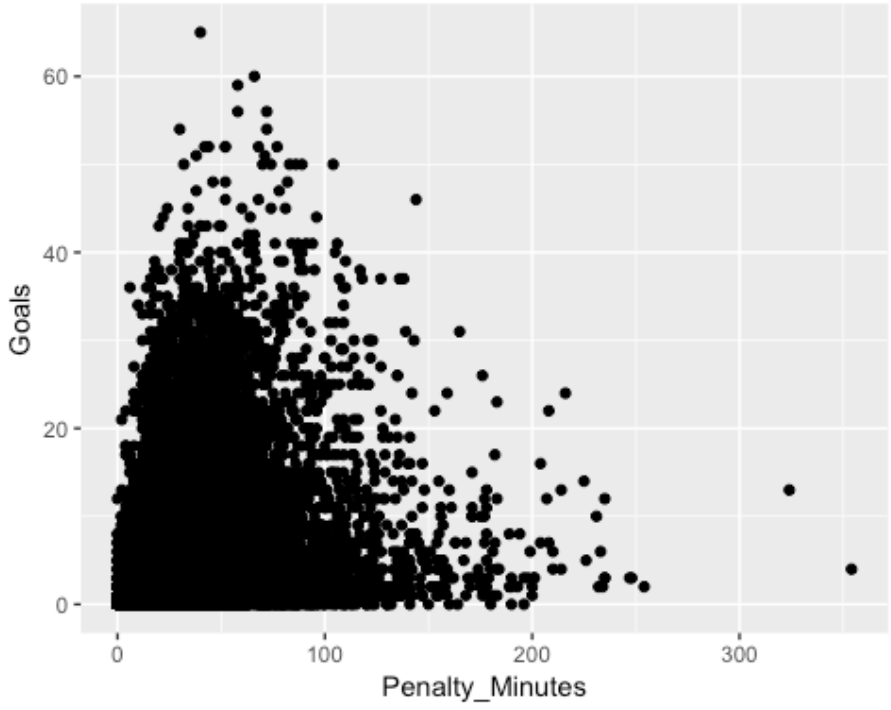
It seems that players drafted younger may score more goals if they were players drafted in early rounds, though there does not appear to be any difference between the teams making and not making the playoffs.

Player Scoring Statistics

For a first look at scoring statistics, we evaluate how shots per game, the number of shots players average for each game, and penalty minutes, the time spent in the penalty box instead of on the ice, impact the rate of goal scoring.

Warning: Removed 23 rows containing missing values (geom_point).

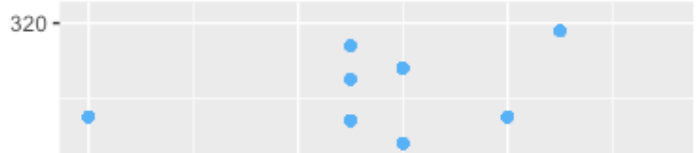


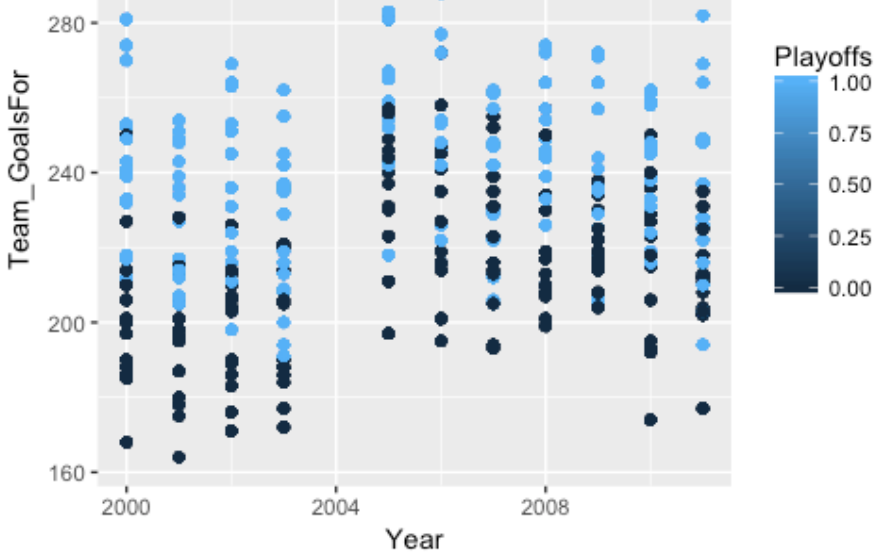


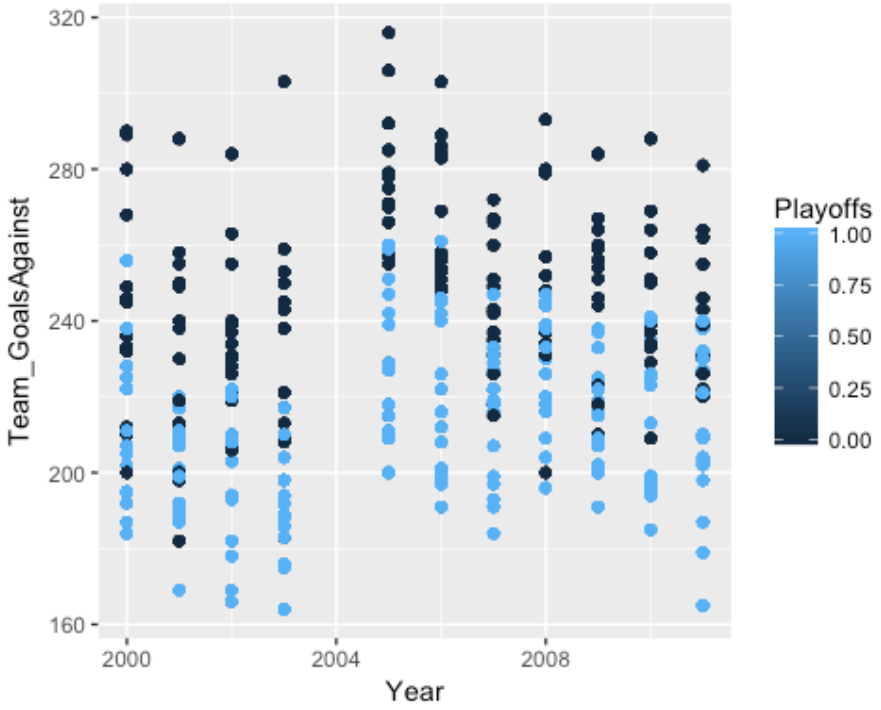
As one might expect, it appears the more shots a player takes, the more goals he scores. The relationship between time spent in the penalty box and goals is less clear but may suggest that more time spent serving penalties leaves less time to score goals.

Team Scoring Statistics

It could be hypothesized that teams who score more goals in a year and allows less goals in a year are more likely to go to the playoffs.

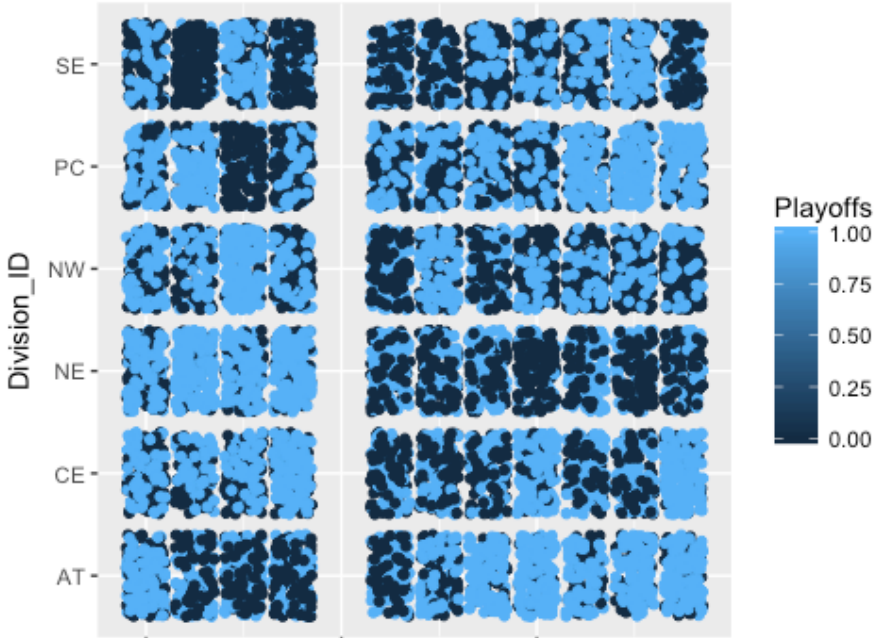


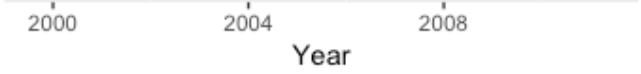




This hypothesis appears to be true for the highest goal-scoring teams and lowest goal-allowing teams year-to-year.

Historically, some teams and divisions (geographical groupings of teams) are consistently successful while others struggle.



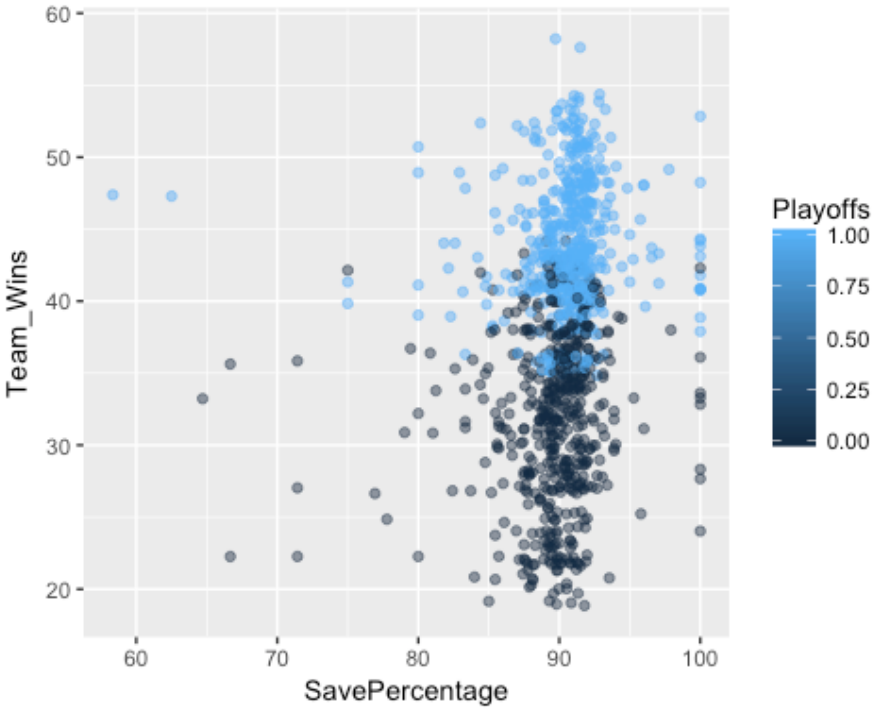


This graph may suggest that the SouthEast (SE) division is more hit-or-miss in terms of making the playoffs than the NorthEast (NE) division prior to the 2004 lockout or the Atlantic (AT) division in the years after the lockout.

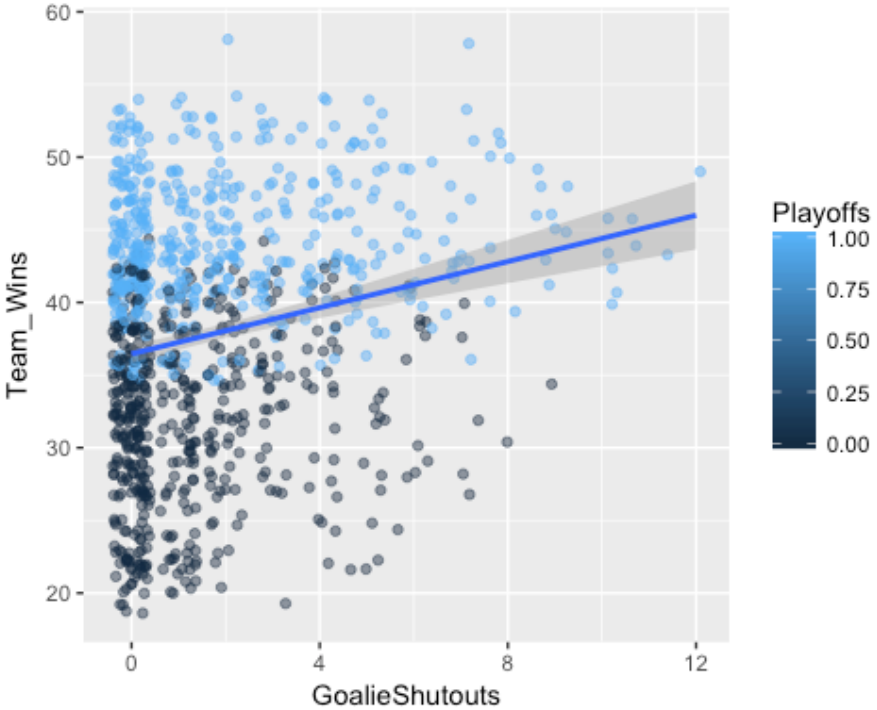
Goaltender Statistics

Another interesting area to evaluate is defensive statistics related to goaltenders. A goaltender's save percentage (the percent of shots they save from becoming goals) and shutouts (games where they don't allow the other team to score at all) should influence their team's win rate and perhaps making the playoffs.

```
ggplot(Hockey, aes(x = SavePercentage, y=Team_Wins, col=Playoffs)) + geom_point(alpha = 0.5, position = "jitter")
```
```
## Warning: Removed 9443 rows containing missing values (geom_point).
```




```
ggplot(Hockey, aes(x = GoalieShutouts, y=Team_Wins, col=Playoffs)) + geom_point(alpha = 0.5, position = "jitter") + geom_smooth(method = lm)
```
```
## Warning: Removed 9440 rows containing non-finite values (stat_smooth).
## Warning: Removed 9440 rows containing missing values (geom_point).
```



From these graphs the relationship is still somewhat unclear.

Summary

This initial statistical and graphical foray into the dataset demonstrates that there are many potential trends to explore further in the model as well as a lot of unclear relationships at this early stage. With an eye towards evaluating how well draft position predicts future player and team success, it may be very interesting to investigate how early in a player's career this impact is made.

Appendix 1

```
## Look at the main dataset.
summary(Hockey)
##    Team_ID             Year       Player_ID          First_Name
##  Length:10442       Min.   :2000   Length:10442       Length:10442
##  Class :character   1st Qu.:2002   Class :character   Class :character
##  Mode  :character   Median :2006   Mode  :character   Mode  :character
##                     Mean   :2005
##                     3rd Qu.:2009
##                     Max.   :2011
##
##   Last_Name            Height          Weight      Shooting_Hand
##  Length:10442       Min.   :66.00   Min.   :144.0   Length:10442
##  Class :character   1st Qu.:72.00   1st Qu.:192.0   Class :character
##  Mode  :character   Median :73.00   Median :203.0   Mode  :character
##                     Mean   :73.27   Mean   :203.4
##                     3rd Qu.:75.00   3rd Qu.:214.0
##                     Max.   :81.00   Max.   :265.0
##
##  FirstNHLseason LastNHLseason  Position_Played    Birth_Country
##  Min.   :1979   Min.   :2000   Length:10442       Length:10442
##  1st Qu.:1995   1st Qu.:2006   Class :character   Class :character
##  Median :2000   Median :2010   Mode  :character   Mode  :character
##  Mean   :2000   Mean   :2008
##  3rd Qu.:2005   3rd Qu.:2011
##  Max.   :2011   Max.   :2011
##
##  Birth_State         Birth_City            Stint        Games_Played
##  Length:10442       Length:10442       Min.   :1.000   Min.   : 0.00
##  Class :character   Class :character   1st Qu.:1.000   1st Qu.:16.00
##  Mode  :character   Mode  :character   Median :1.000   Median :50.00
##                                        Mean   :1.094   Mean   :44.93
##                                        3rd Qu.:1.000   3rd Qu.:73.00
##                                        Max.   :3.000   Max.   :82.00
##
##      Goals           Assists          Points       Penalty_Minutes
##  Min.   : 0.000   Min.   : 0.00   Min.   :  0.00   Min.   :  0.00
##  1st Qu.: 0.000   1st Qu.: 1.00   1st Qu.:  1.00   1st Qu.:  5.00
##  Median : 3.000   Median : 6.00   Median : 10.00   Median : 22.00
##  Mean   : 6.618   Mean   :11.32   Mean   : 17.94   Mean   : 32.19
##  3rd Qu.:10.000   3rd Qu.:18.00   3rd Qu.: 28.00   3rd Qu.: 47.00
##  Max.   :65.000   Max.   :92.00   Max.   :123.00   Max.   :354.00
##
##    Plus_Minus          PP_Goals       PP_Assists        SH_Goals
##  Min.   :-46.0000   Min.   : 0.00   Min.   : 0.000   Min.   :0.0000
##  1st Qu.: -4.0000   1st Qu.: 0.00   1st Qu.: 0.000   1st Qu.:0.0000
##  Median :  0.0000   Median : 0.00   Median : 0.000   Median :0.0000
##  Mean   : -0.2552   Mean   : 1.81   Mean   : 3.437   Mean   :0.2297
##  3rd Qu.:  3.0000   3rd Qu.: 2.00   3rd Qu.: 5.000   3rd Qu.:0.0000
##  Max.   : 52.0000   Max.   :27.00   Max.   :48.000   Max.   :9.0000
##
##    SH_Assists     Gamewinning_Goals GameTying_Goals       Shots
##  Min.   :0.0000   Min.   : 0.000    Min.   :0.00000   Min.   :  0.00
##  1st Qu.:0.0000   1st Qu.: 0.000    1st Qu.:0.00000   1st Qu.:  8.00
##  Median :0.0000   Median : 0.000    Median :0.00000   Median : 47.00
##  Mean   :0.2487   Mean   : 1.046    Mean   :0.05545   Mean   : 69.74
##  3rd Qu.:0.0000   3rd Qu.: 1.000    3rd Qu.:0.00000   3rd Qu.:112.00
##  Max.   :7.0000   Max.   :12.000    Max.   :3.00000   Max.   :528.00
##
##  GoalieGames_Played GoalieMinutes      GoalieWins     GoalieLosses
##  Min.   : 1.00      Min.   :   0.0   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 7.00      1st Qu.: 356.5   1st Qu.: 2.00   1st Qu.: 3.00
##  Median :21.00      Median :1144.0   Median : 8.00   Median : 8.00
##  Mean   :27.56      Mean   :1544.0   Mean   :12.08   Mean   :10.35
##  3rd Qu.:45.00      3rd Qu.:2493.0   3rd Qu.:20.00   3rd Qu.:17.00
```

```
##  Max.   :78.00      Max.   :4697.0   Max.   :48.00   Max.   :41.00
##  NA's   :9440       NA's   :9440     NA's   :9440    NA's   :9440
##  GoalieShutouts GoalieGoals_Against GoalieShots_Against   Draft_Pick
##  Min.   : 0.0   Min.   :  0.00      Min.   :   0.0      Min.   :  1.00
##  1st Qu.: 0.0   1st Qu.: 18.00      1st Qu.: 167.2      1st Qu.: 21.00
##  Median : 1.0   Median : 53.00      Median : 568.5      Median : 59.00
##  Mean   : 1.7   Mean   : 68.32      Mean   : 742.7      Mean   : 83.05
##  3rd Qu.: 3.0   3rd Qu.:114.00      3rd Qu.:1216.0      3rd Qu.:133.00
##  Max.   :12.0   Max.   :232.00      Max.   :2488.0      Max.   :291.00
##  NA's   :9440   NA's   :9440        NA's   :9440
##    Draft_Year    Draft_Team          Draft_Age     AmateurTeam
##  Min.   :1979   Length:10442       Min.   :17.00   Length:10442
##  1st Qu.:1993   Class :character   1st Qu.:18.00   Class :character
##  Median :1998   Mode  :character   Median :18.00   Mode  :character
##  Mean   :1997                      Mean   :18.62
##  3rd Qu.:2002                      3rd Qu.:19.00
##  Max.   :2010                      Max.   :31.00
##
##  AmateurLeague        BirthDate            DeathDate
##  Length:10442       Min.   :1960-12-03   Min.   :2003-07-05
##  Class :character   1st Qu.:1975-01-15   1st Qu.:2011-08-31
##  Mode  :character   Median :1979-02-15   Median :2011-09-07
##                     Mean   :1979-01-11   Mean   :2010-09-27
##                     3rd Qu.:1983-06-23   3rd Qu.:2011-09-07
##                     Max.   :1992-09-08   Max.   :2011-09-07
##                                          NA's   :10352
##  Conference_ID      Division_ID        SeasonEnd_Rank  Playoff_Result
##  Length:10442       Length:10442       Min.   :1.000   Length:10442
##  Class :character   Class :character   1st Qu.:2.000   Class :character
##  Mode  :character   Mode  :character   Median :3.000   Mode  :character
##                                        Mean   :3.062
##                                        3rd Qu.:4.000
##                                        Max.   :5.000
##
##  Team_Total_Games   Team_Wins      Team_Losses      Team_Ties
##  Min.   :82       Min.   :19.00   Min.   :15.00   Min.   : 0.00
##  1st Qu.:82       1st Qu.:33.00   1st Qu.:27.00   1st Qu.: 0.00
##  Median :82       Median :40.00   Median :31.00   Median : 0.00
##  Mean   :82       Mean   :38.55   Mean   :31.73   Mean   : 4.03
##  3rd Qu.:82       3rd Qu.:44.00   3rd Qu.:36.00   3rd Qu.: 9.00
##  Max.   :82       Max.   :58.00   Max.   :51.00   Max.   :20.00
##
##  Team_OT_Losses    Team_Points     Team_ShootoutWins Team_ShootoutLosses
##  Min.   : 0.000   Min.   : 52.00   Min.   : 0.000    Min.   : 0.000
##  1st Qu.: 5.000   1st Qu.: 78.00   1st Qu.: 0.000    1st Qu.: 0.000
##  Median : 8.000   Median : 91.00   Median : 3.000    Median : 3.000
##  Mean   : 7.696   Mean   : 88.82   Mean   : 3.302    Mean   : 3.345
##  3rd Qu.:10.000   3rd Qu.:100.00   3rd Qu.: 6.000    3rd Qu.: 6.000
##  Max.   :18.000   Max.   :124.00   Max.   :15.000    Max.   :12.000
##
##  Team_GoalsFor   Team_GoalsAgainst   TeamName         Team_PenaltyMin
##  Min.   :164.0   Min.   :164       Length:10442       Min.   : 689
##  1st Qu.:209.0   1st Qu.:208       Class :character   1st Qu.:1020
##  Median :227.0   Median :229       Mode  :character   Median :1155
##  Mean   :227.7   Mean   :230                          Mean   :1159
##  3rd Qu.:247.0   3rd Qu.:250                          3rd Qu.:1287
##  Max.   :318.0   Max.   :316                          Max.   :1994
##
##  Team_BenchMinors    Team_PPG         Playoffs       Draft_Round
##  Min.   : 4.00    Min.   : 33.00   Min.   :0.0000   Min.   :1.000
##  1st Qu.:14.00    1st Qu.: 50.00   1st Qu.:0.0000   1st Qu.:1.000
##  Median :18.00    Median : 58.00   Median :1.0000   Median :2.000
##  Mean   :18.41    Mean   : 60.67   Mean   :0.5106   Mean   :3.244
##  3rd Qu.:22.00    3rd Qu.: 70.00   3rd Qu.:1.0000   3rd Qu.:5.000
##  Max.   :50.00    Max.   :107.00   Max.   :1.0000   Max.   :8.000
##
##   GoalsPerGame     PointsPerGame      ShotsPerGame     PercentGoals
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   : 0.000
##  1st Qu.:0.00000   1st Qu.:0.08571   1st Qu.:0.6555   1st Qu.: 0.000
##  Median :0.07407   Median :0.25000   Median :1.1875   Median : 1.232
##  Mean   :0.11330   Mean   :0.31241   Mean   :1.2883   Mean   : 2.892
##  3rd Qu.:0.17732   3rd Qu.:0.47059   3rd Qu.:1.8289   3rd Qu.: 4.292
##  Max.   :1.00000   Max.   :1.76744   Max.   :6.6835   Max.   :29.500
##  NA's   :23        NA's   :23        NA's   :23
##   PercentGames    YearsExperience SavePercentage   BirthRegion
```

```
##  Min.   :  0.00   Min.   : 0.00   Min.   : 58.33   Length:10442
##  1st Qu.: 19.51   1st Qu.: 2.00   1st Qu.: 89.17   Class :character
##  Median : 60.98   Median : 5.00   Median : 90.53   Mode  :character
##  Mean   : 54.79   Mean   : 5.77   Mean   : 90.07
##  3rd Qu.: 89.02   3rd Qu.: 9.00   3rd Qu.: 91.62
##  Max.   :100.00   Max.   :26.00   Max.   :100.00
##                                   NA's   :9443
```

Appendix 2

Draft Round by Country

Birth_Country	MeanDraftRound
Austria	4.0
Belarus	2.5
Brazil	1.0
Brunei Darussalam	8.0
Canada	3.3
Croatia	8.0
Czech Republic	3.6
Denmark	3.5
England	5.0
Finland	4.3
France	8.0
Germany	3.8
Ireland	1.0
Italy	1.0
Japan	4.5
Kazakhstan	4.3
Latvia	4.5
Lithuania	1.0
Norway	5.2
Poland	2.7
Russia	3.2
Scotland	4.0
Slovakia	4.0
Slovenia	1.0
South Africa	1.0
South Korea	2.0
Sweden	3.8
Switzerland	4.8
Tanzania	2.0
Ukraine	2.8
USA	3.6
USSR	2.0
Yugoslavia	7.0
NA	3.2

Draft Age by Country

Birth_Country	MeanDraftAge
Austria	21.6
Belarus	18.8
Brazil	18.0
Brunei Darussalam	19.0
Canada	18.3
Croatia	19.0
Czech Republic	19.3
Denmark	18.0
England	18.0
Finland	20.4

France	26.0
Germany	18.6
Ireland	18.0
Italy	18.0
Japan	19.5
Kazakhstan	18.1
Latvia	19.8
Lithuania	18.5
Norway	18.8
Poland	19.0
Russia	19.4
Scotland	18.0
Slovakia	18.5
Slovenia	18.0
South Africa	19.0
South Korea	18.0
Sweden	19.2
Switzerland	20.2
Tanzania	18.0
Ukraine	18.6
USA	18.6
USSR	18.0
Yugoslavia	18.0
NA	18.3