Final Project

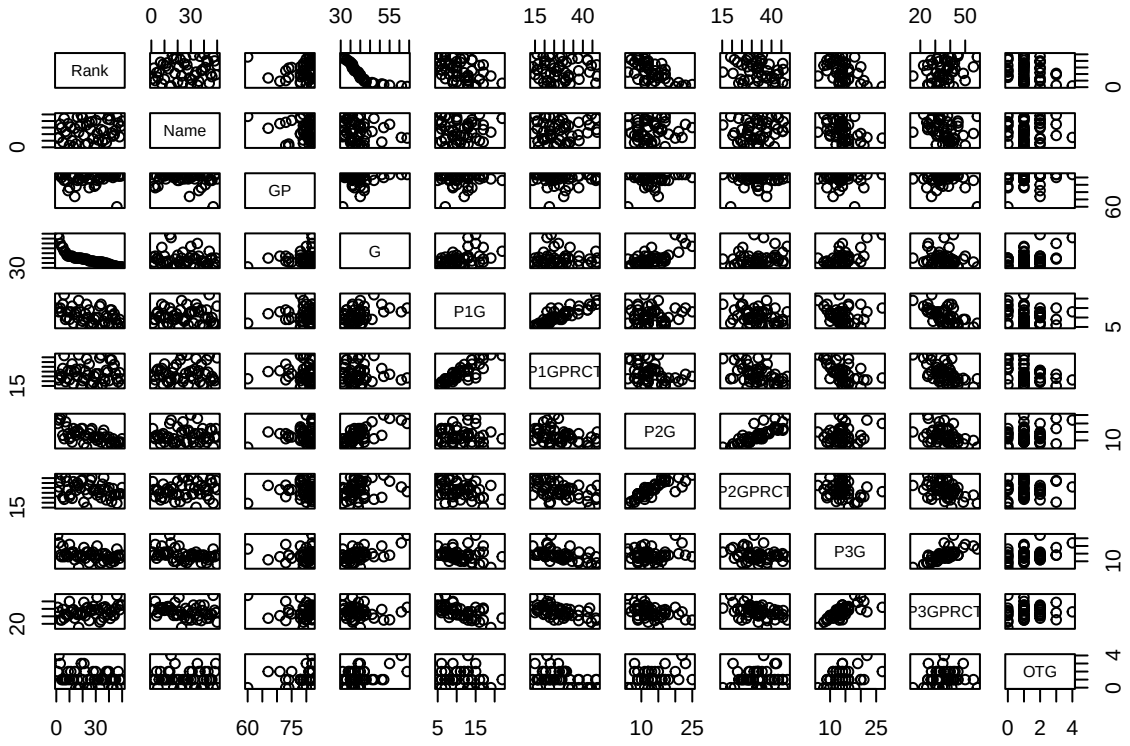
Scott Saxton

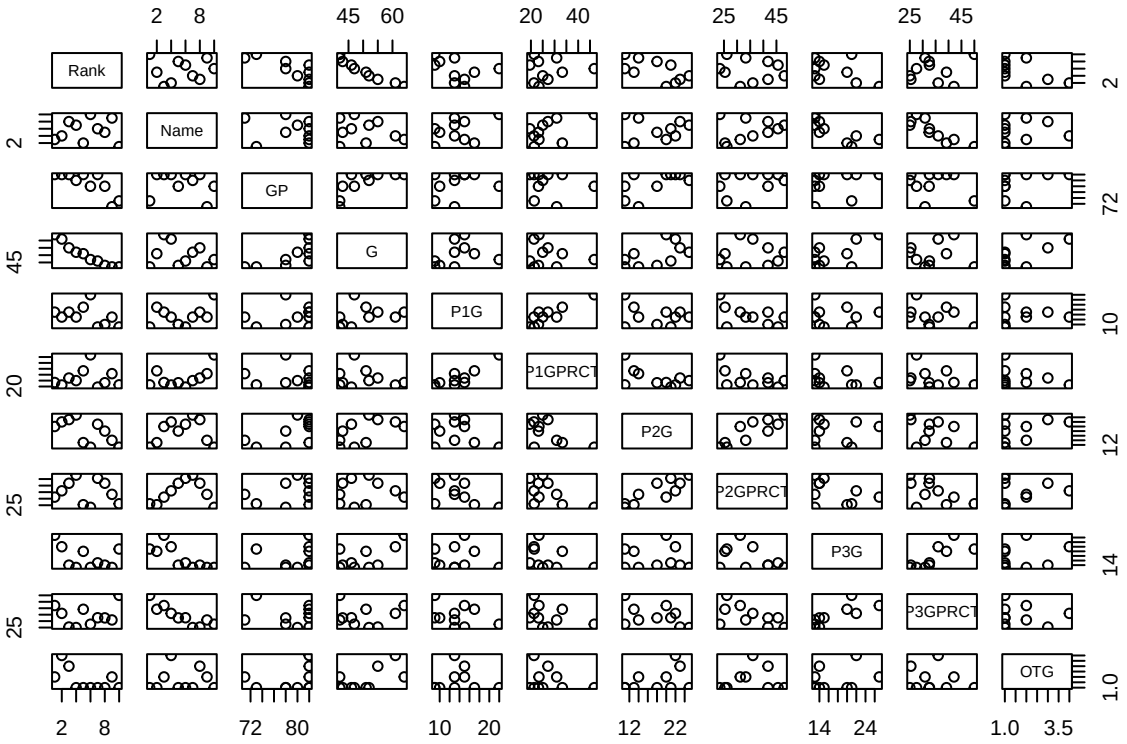
2023-12-08

The basis of this project was to explore if the National Hockey League's (NHL) top 50 goal scorers during the 2022-2023 regular season had a tendency to score more goals during one specific period versus other periods. The data was collected from quanthockey.com and supplemental data from hockey-reference.com. The source for goals per period required the data be extracted from each team's box score for all 82 games. Once the basic data was collected and organized, the players were ranked from 1 to 50 based on their total goals scored for the season. Next, each players goals were categorized by the period it was scored. Then the percentage of their per period goals was calculated against their total season goals. The percentage was rounded to 1 tenth to simplify the data set. No outliers were removed from the data set because all values were pertinent to the overall data. The data set includes overtime goals in the total goals for the season, this will account for the small variation in the percentages of goals per period. Additionally the number of games played was tracked but not used to adjust the percentages. Finally, the data set was summarized in subset, Overall Statistics, First through Third Period Statistics and Final Findings.

The Overall Statistics showed that most players fell within the 1st(35) and 3rd(40.75) quartile, with only 5 outliers in the data set.
The Mean was 39.04 goals and the Median was 37.00 goals.
The data set follows a linear projection.
The following Data visualizations show the data set, full data set plot, subset of top 10 players, data summary, scatterplot of ranking vs goals and boxplot of ranking vs goals.
A total of 1952 goals were scored by the top 50 ranked players.
The summary of the period totals versus the overall total will be covered at the end of this project.

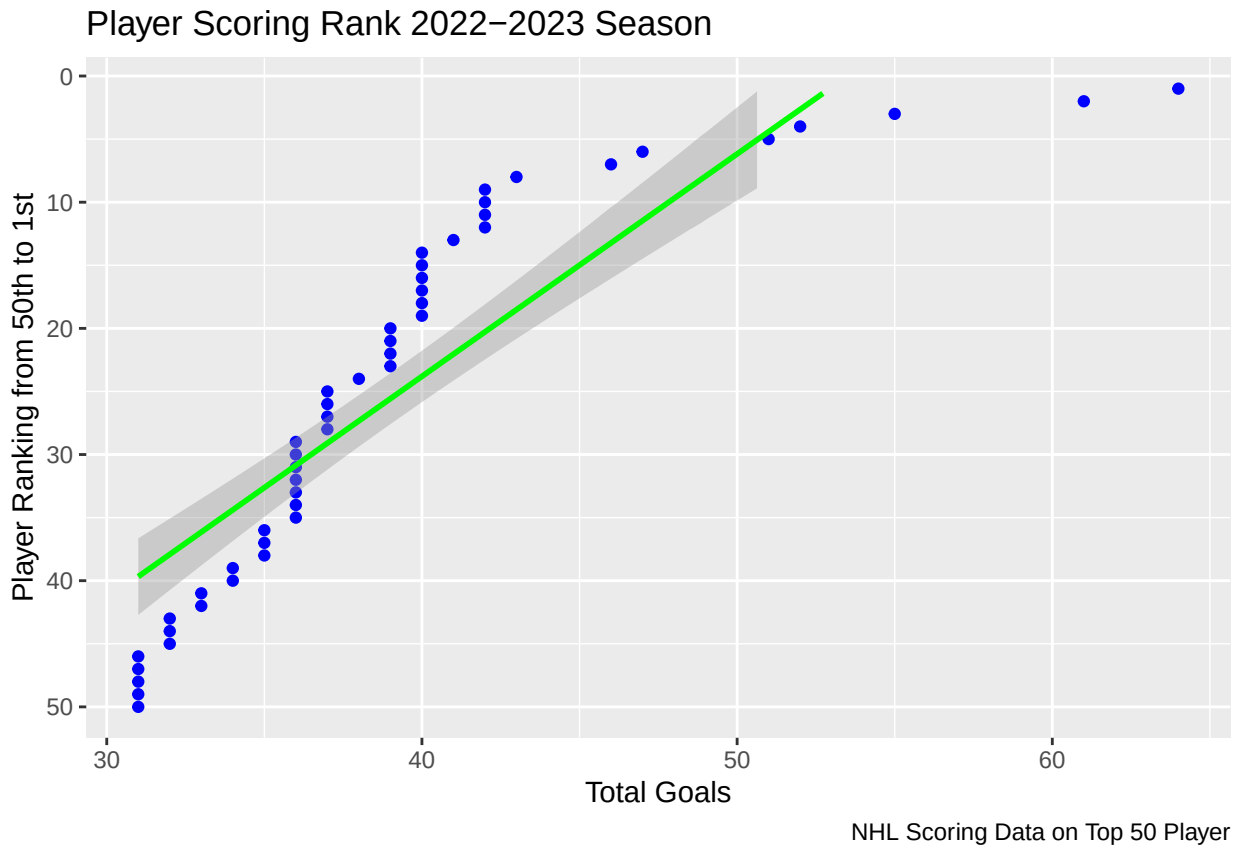
```
##   Rank           Name GP  G P1G P1GPRCT P2G P2GPRCT P3G P3GPRCT OTG
## 1   50   Sam Reinhart 82 31   5    16.1  11    35.4  14    45.1   1
## 2   49  Matthew Boldy 81 31   7    22.5   9    29.0  14    45.1   1
## 3   48   Dylan Cozens 81 31   9    29.0   9    29.0  12    38.7   1
## 4   47  Nico Hischier 81 31   6    19.3   9    29.0  14    45.1   2
## 5   46 Travis Konecny 60 31   7    22.5   6    19.3  18    58.1   0
## 6   45   Jesper Bratt 82 32   9    28.1   9    28.1  13    40.6   1
```


Rank
Name
GP
G
P1G
P1GPRCT
P2G
P2GPRCT
P3G
P3GPRCT
OTG

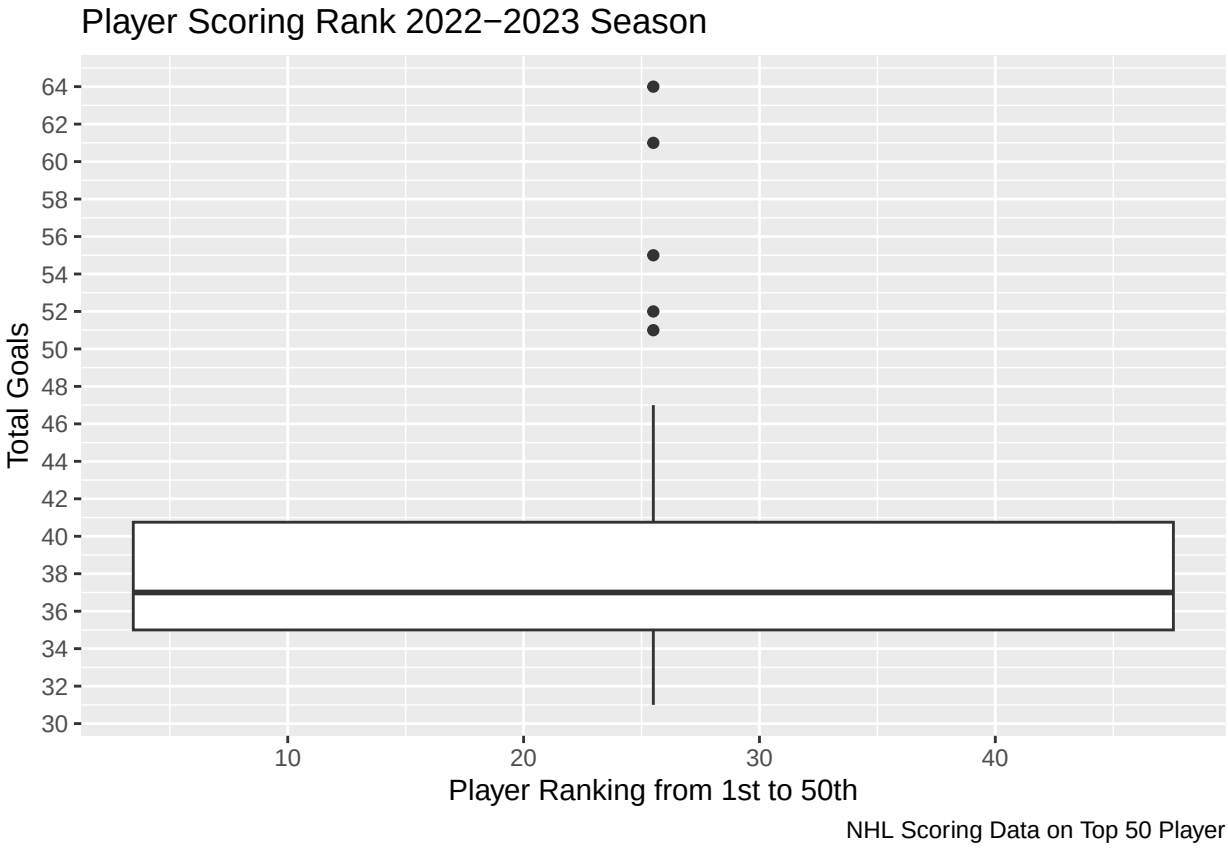



```
##      Name                GP              G              P1G
##  Length:50          Min.   :60.00   Min.   :31.00   Min.   : 5.00
##  Class :character   1st Qu.:78.25   1st Qu.:35.00   1st Qu.: 7.25
##  Mode  :character   Median :81.00   Median :37.00   Median :10.00
##                     Mean   :79.08   Mean   :39.04   Mean   :10.88
##                     3rd Qu.:82.00   3rd Qu.:40.75   3rd Qu.:13.00
##                     Max.   :82.00   Max.   :64.00   Max.   :22.00
##     P1GPRCT           P2G            P2GPRCT           P3G        P3GPRCT
##  Min.   :13.80   Min.   : 6.00   Min.   :15.30   Min.   : 6   Min.   :15.00
##  1st Qu.:20.70   1st Qu.: 9.25   1st Qu.:28.10   1st Qu.:12   1st Qu.:31.20
##  Median :26.75   Median :12.00   Median :33.30   Median :14   Median :36.15
##  Mean   :27.88   Mean   :12.96   Mean   :32.91   Mean   :14   Mean   :36.12
##  3rd Qu.:32.50   3rd Qu.:15.75   3rd Qu.:37.73   3rd Qu.:15   3rd Qu.:40.90
##  Max.   :47.50   Max.   :25.00   Max.   :48.00   Max.   :27   Max.   :58.10
##       OTG
##  Min.   :0.0
##  1st Qu.:1.0
##  Median :1.0
##  Mean   :1.2
##  3rd Qu.:2.0
##  Max.   :4.0
```

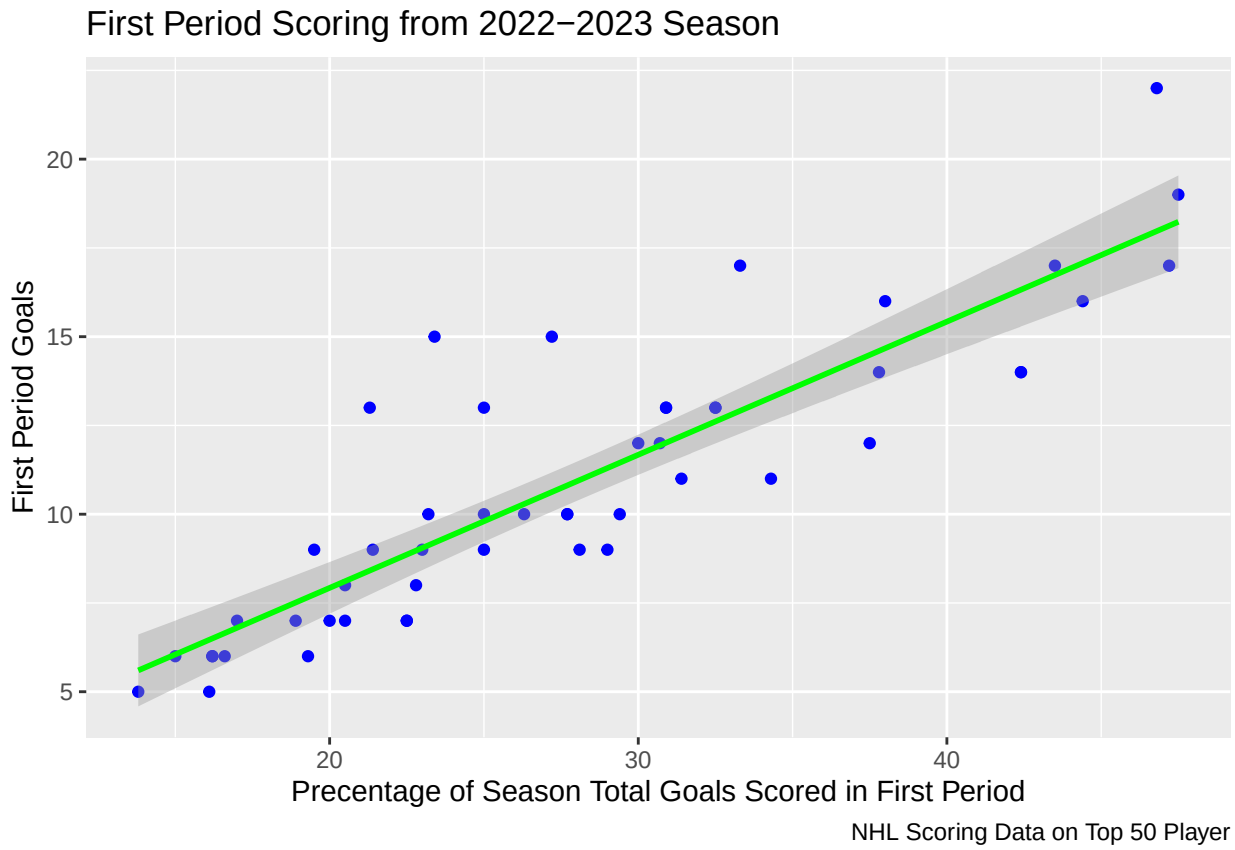
```
## Warning: Removed 27 rows containing missing values (‘geom_smooth()‘).
```



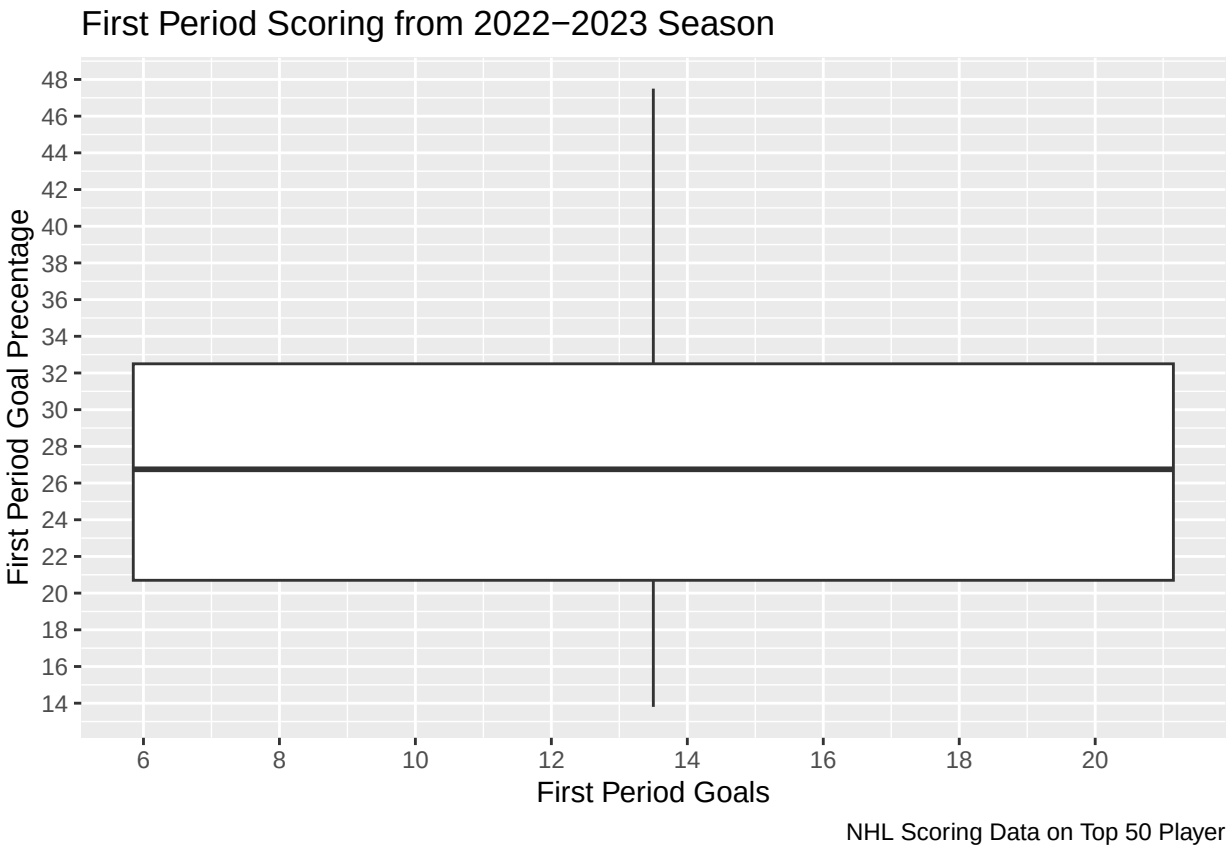

```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```

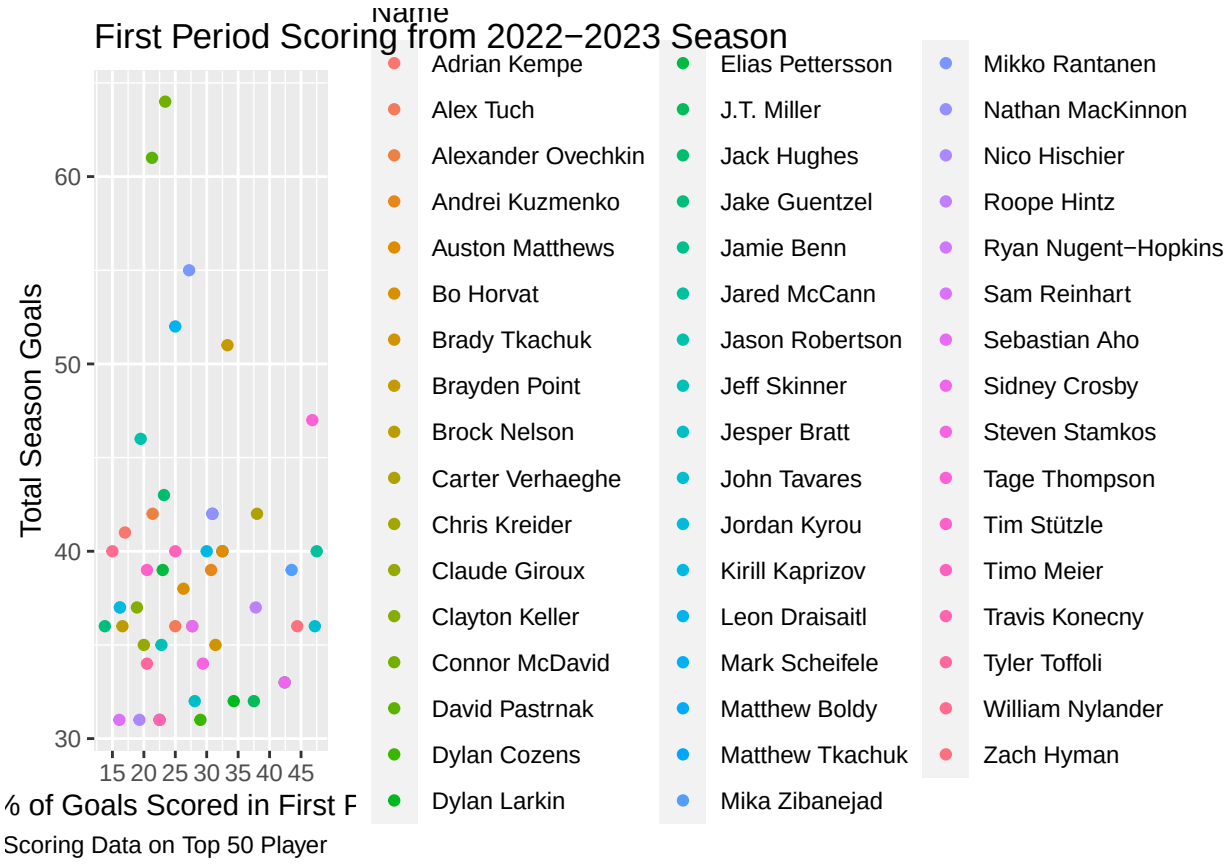

Player Scoring Rank 2022−2023 Season
Total Goals
64
62
60
58
56
54
52
50
48
46
44
42
40
38
36
34
32
30
10
20
30
40
Player Ranking from 1st to 50th
NHL Scoring Data on Top 50 Player

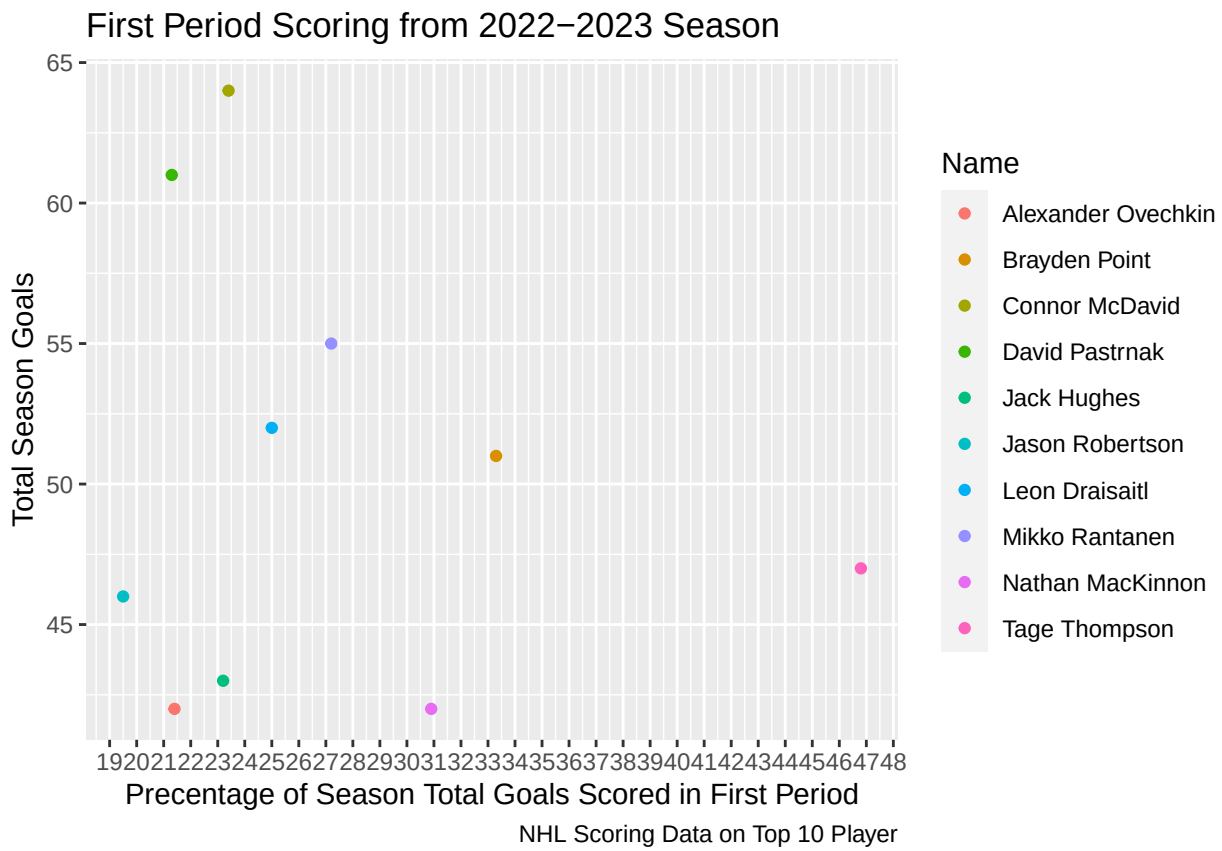
The first period data shows that the data following a linear projection with few outliers.
The first period Statistics showed that most players fell within the 1st(7.25) and 3rd(13.00) quartile, with only 5 outliers in the data set.
The Mean was 10.88 goals and the Median was 10.00 goals.
The one player that had the highest percentage of first period goals was Jared McCann (47.5).
The player with the most first period goals was Tage Thompson, whom is a top 10 scorer (22).
The Top 10 Scores did not follow a linear projection in percentage, with a range from 19.5 to 46.8 percent.
All but two of the Top 10 Scores were below 33%, indicating the first period on average had less goals scored.




```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```

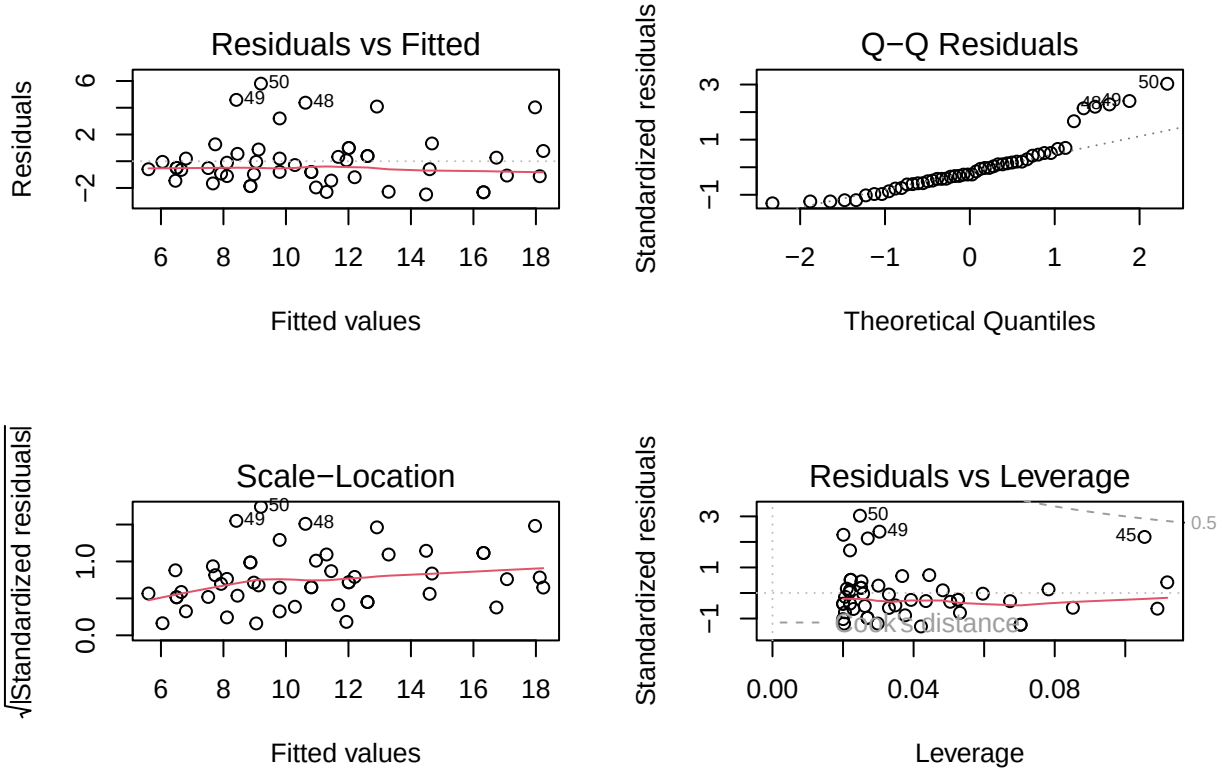

First Period Scoring from 2022−2023 Season
First Period Goal Precentage
48
46
44
42
40
38
36
34
32
30
28
26
24
22
20
18
16
14
6
8
10
12
14
16
18
20
First Period Goals
NHL Scoring Data on Top 50 Player


First Period Scoring from 2022−2023 Season
Name
Adrian Kempe
Alex Tuch
Alexander Ovechkin
Andrei Kuzmenko
Auston Matthews
Bo Horvat
Brady Tkachuk
Brayden Point
Brock Nelson
Carter Verhaeghe
Chris Kreider
Claude Giroux
Clayton Keller
Connor McDavid
David Pastrnak
Dylan Cozens
Dylan Larkin
Elias Pettersson
J.T. Miller
Jack Hughes
Jake Guentzel
Jamie Benn
Jared McCann
Jason Robertson
Jeff Skinner
Jesper Bratt
John Tavares
Jordan Kyrou
Kirill Kaprizov
Leon Draisaitl
Mark Scheifele
Matthew Boldy
Matthew Tkachuk
Mika Zibanejad
Mikko Rantanen
Nathan MacKinnon
Nico Hischier
Roope Hintz
Ryan Nugent−Hopkins
Sam Reinhart
Sebastian Aho
Sidney Crosby
Steven Stamkos
Tage Thompson
Tim Stützle
Timo Meier
Travis Konecny
Tyler Toffoli
William Nylander
Zach Hyman
Total Season Goals
60
50
40
30
15 20 25 30 35 40 45
% of Goals Scored in First P
Scoring Data on Top 50 Player

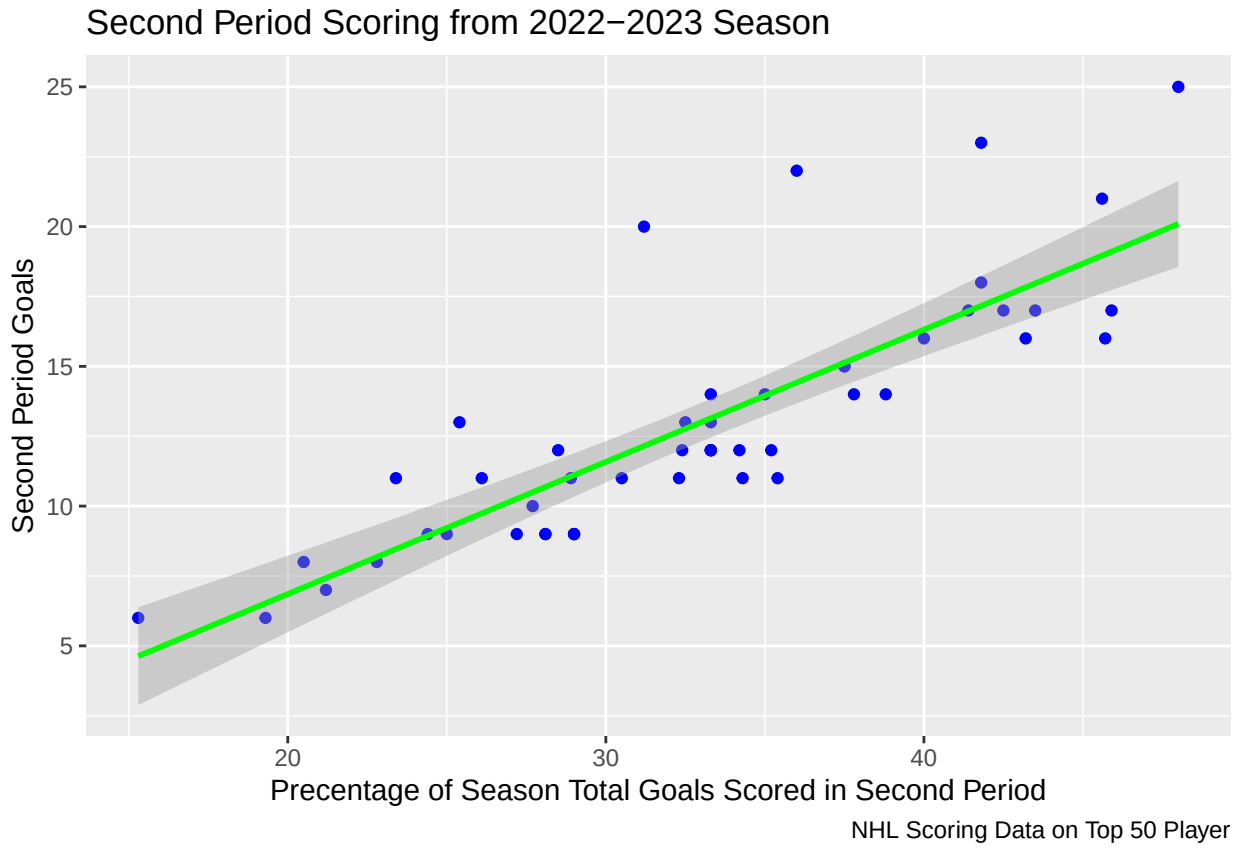



```
## [1] 0.8747257
```

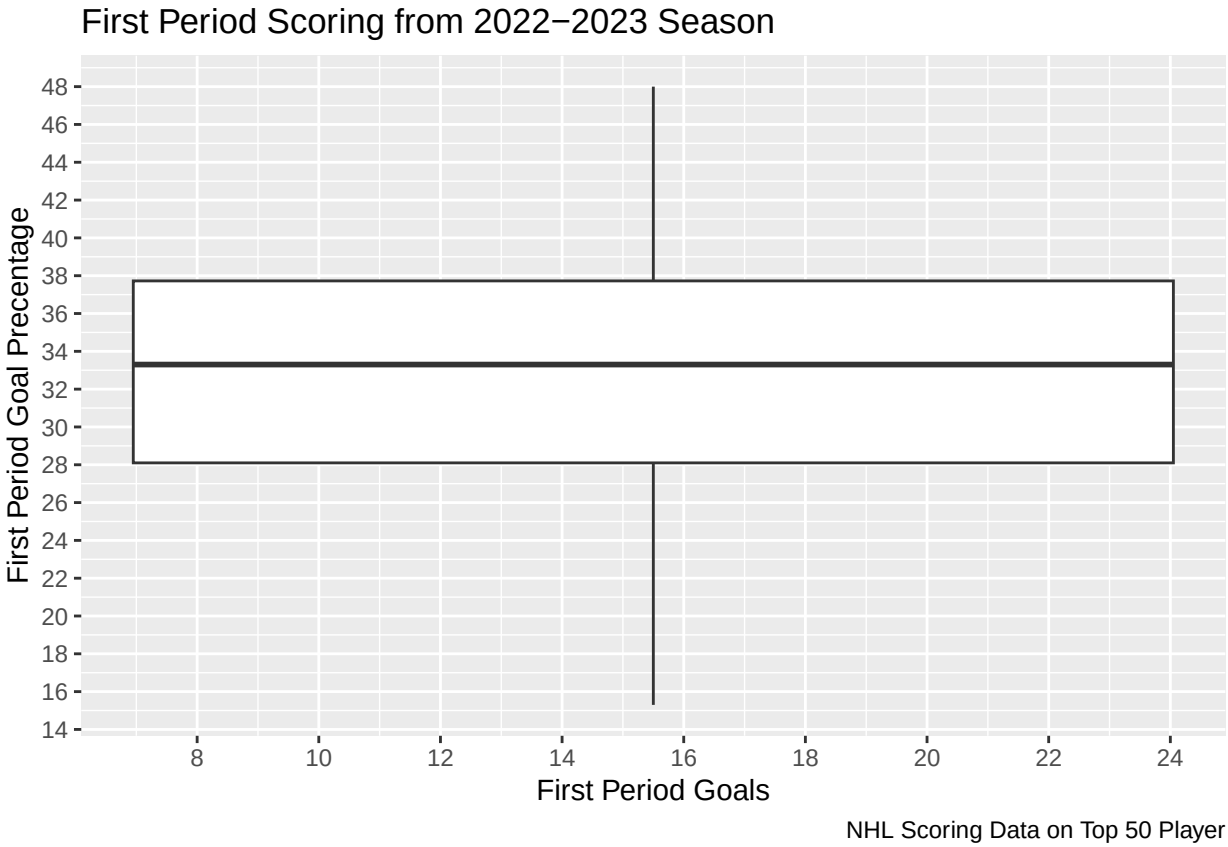
```
##
## Call:
## lm(formula = P1G ~ P1GPRCT, data = Goals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4859 -1.1200 -0.5003  0.5098  5.8004
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.42667    0.87979   0.485     0.63
## P1GPRCT      0.37491    0.02998  12.505   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.94 on 48 degrees of freedom
## Multiple R-squared:  0.7651, Adjusted R-squared:  0.7603
## F-statistic: 156.4 on 1 and 48 DF,  p-value: < 2.2e-16
```

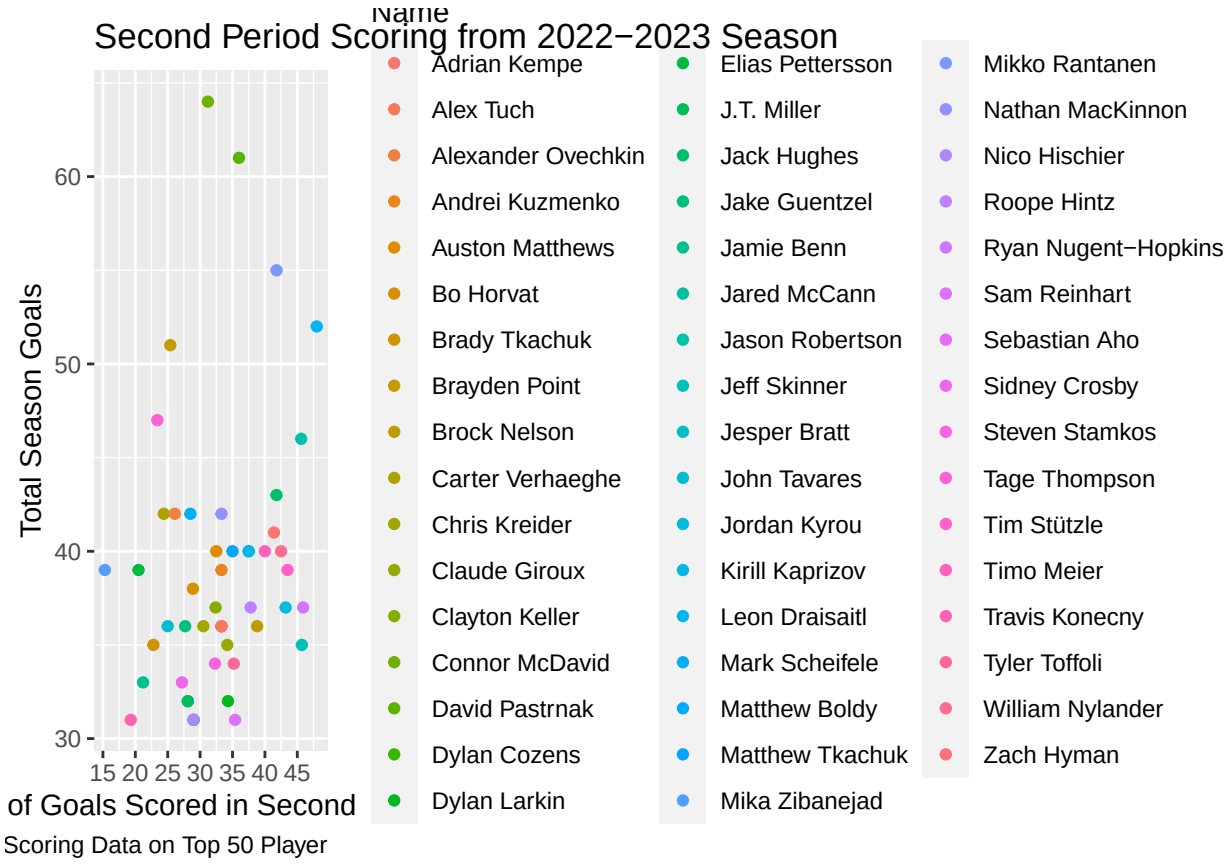

Residuals vs Fitted
Residuals
Fitted values
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance

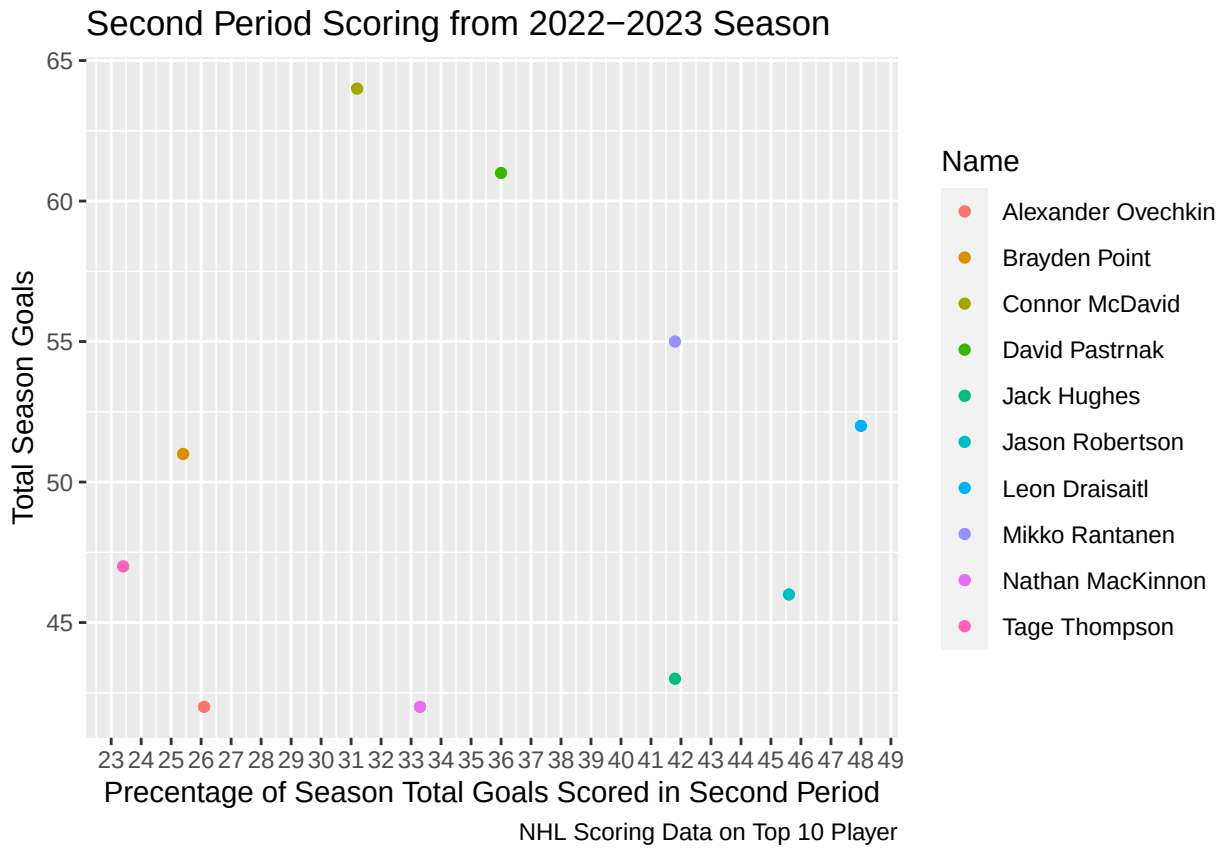
The second period data shows that the data following a linear projection with few outliers.
The second period Statistics showed that most players fell within the 1st(9.25) and 3rd(15.75) quartile, with only 4 outliers in the data set.
The Mean was 12.96 goals and the Median was 12.00 goals.
The one player that had the highest percentage of second period goals was Leon Draisaitl (48.0).
The player with the most second period goals was again Leon Draisaitl, whom is a top 10 scorer (25).
The Top 10 Scores did not follow a linear projection in percentage, with a range from 23.4 to 48.0 percent.
All but four of the Top 10 Scores were above 33%, indicating the Second Period on average had more goals than the First Period.




```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```

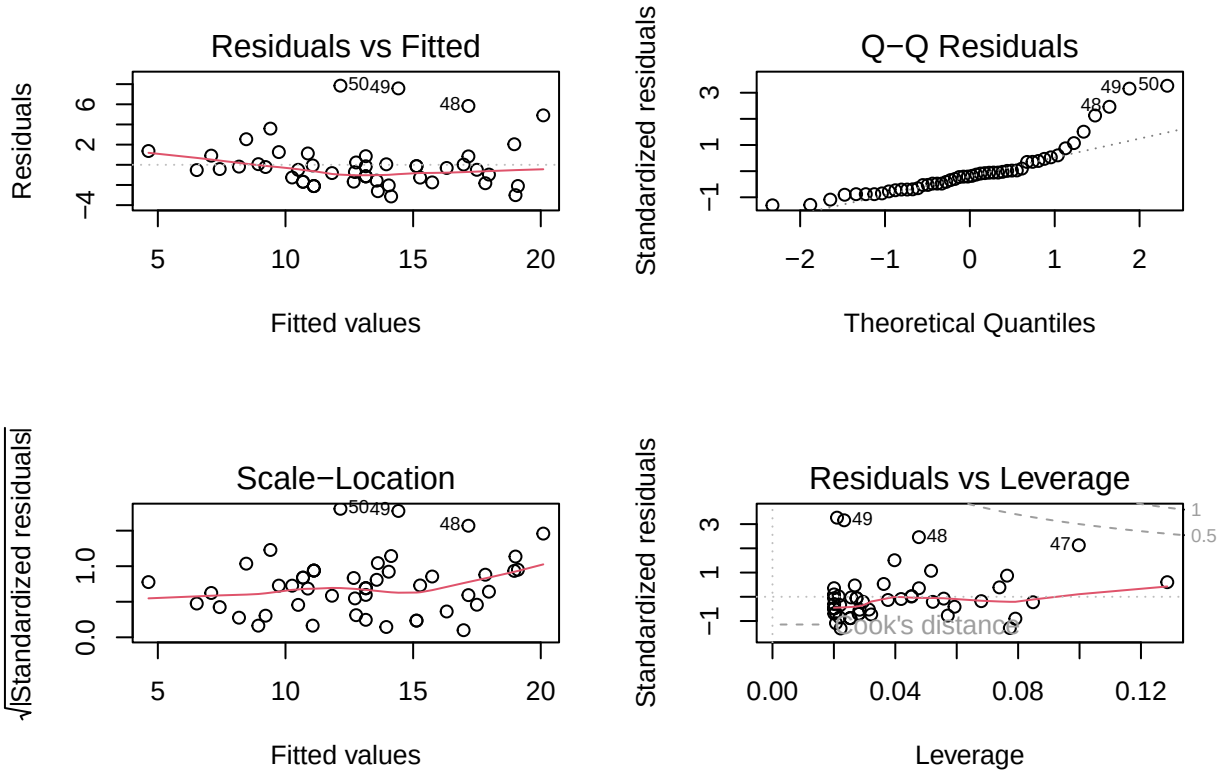

First Period Scoring from 2022−2023 Season
First Period Goal Precentage
48
46
44
42
40
38
36
34
32
30
28
26
24
22
20
18
16
14
8
10
12
14
16
18
20
22
24
First Period Goals
NHL Scoring Data on Top 50 Player


Second Period Scoring from 2022−2023 Season
Total Season Goals
60
50
40
30
15 20 25 30 35 40 45
of Goals Scored in Second
Scoring Data on Top 50 Player
Name
Adrian Kempe
Alex Tuch
Alexander Ovechkin
Andrei Kuzmenko
Auston Matthews
Bo Horvat
Brady Tkachuk
Brayden Point
Brock Nelson
Carter Verhaeghe
Chris Kreider
Claude Giroux
Clayton Keller
Connor McDavid
David Pastrnak
Dylan Cozens
Dylan Larkin
Elias Pettersson
J.T. Miller
Jack Hughes
Jake Guentzel
Jamie Benn
Jared McCann
Jason Robertson
Jeff Skinner
Jesper Bratt
John Tavares
Jordan Kyrou
Kirill Kaprizov
Leon Draisaitl
Mark Scheifele
Matthew Boldy
Matthew Tkachuk
Mika Zibanejad
Mikko Rantanen
Nathan MacKinnon
Nico Hischier
Roope Hintz
Ryan Nugent−Hopkins
Sam Reinhart
Sebastian Aho
Sidney Crosby
Steven Stamkos
Tage Thompson
Tim Stützle
Timo Meier
Travis Konecny
Tyler Toffoli
William Nylander
Zach Hyman

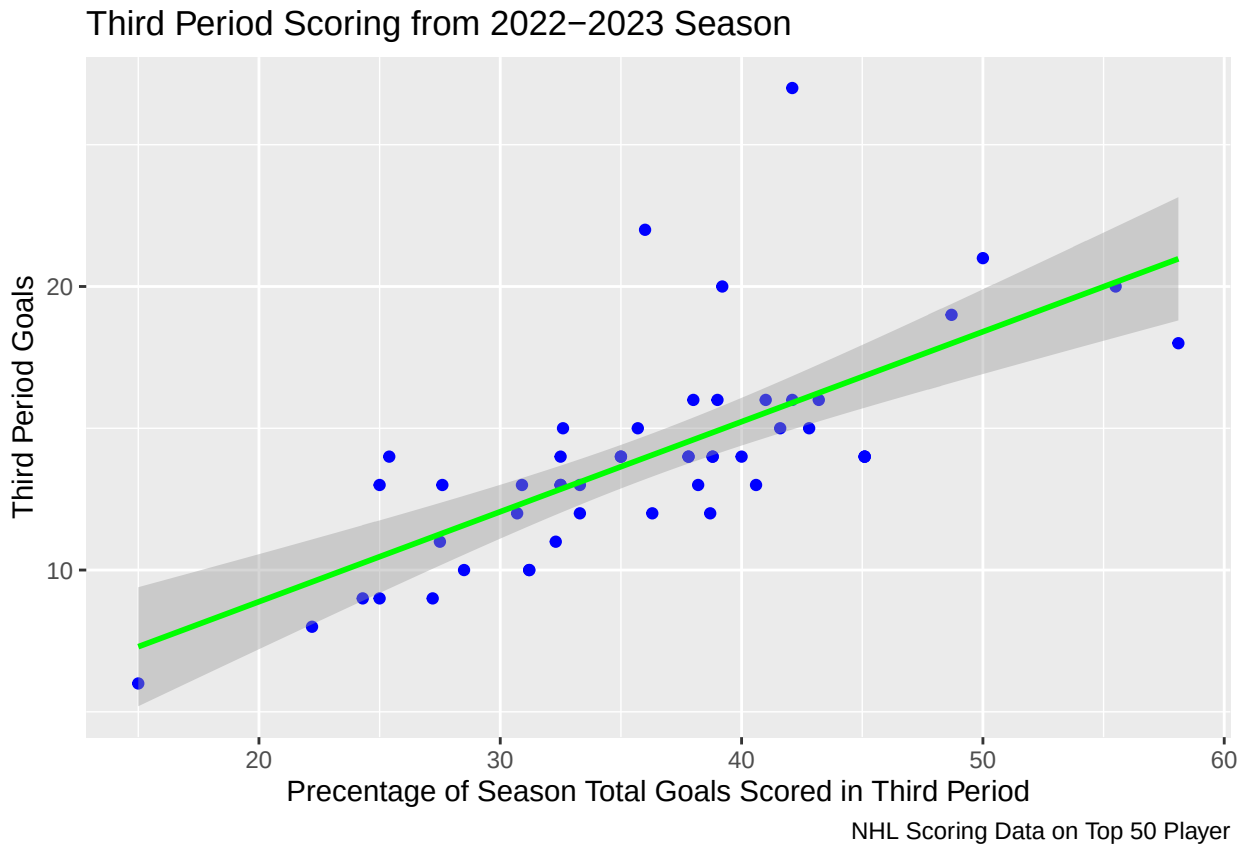



```
## [1] 0.8323829
```

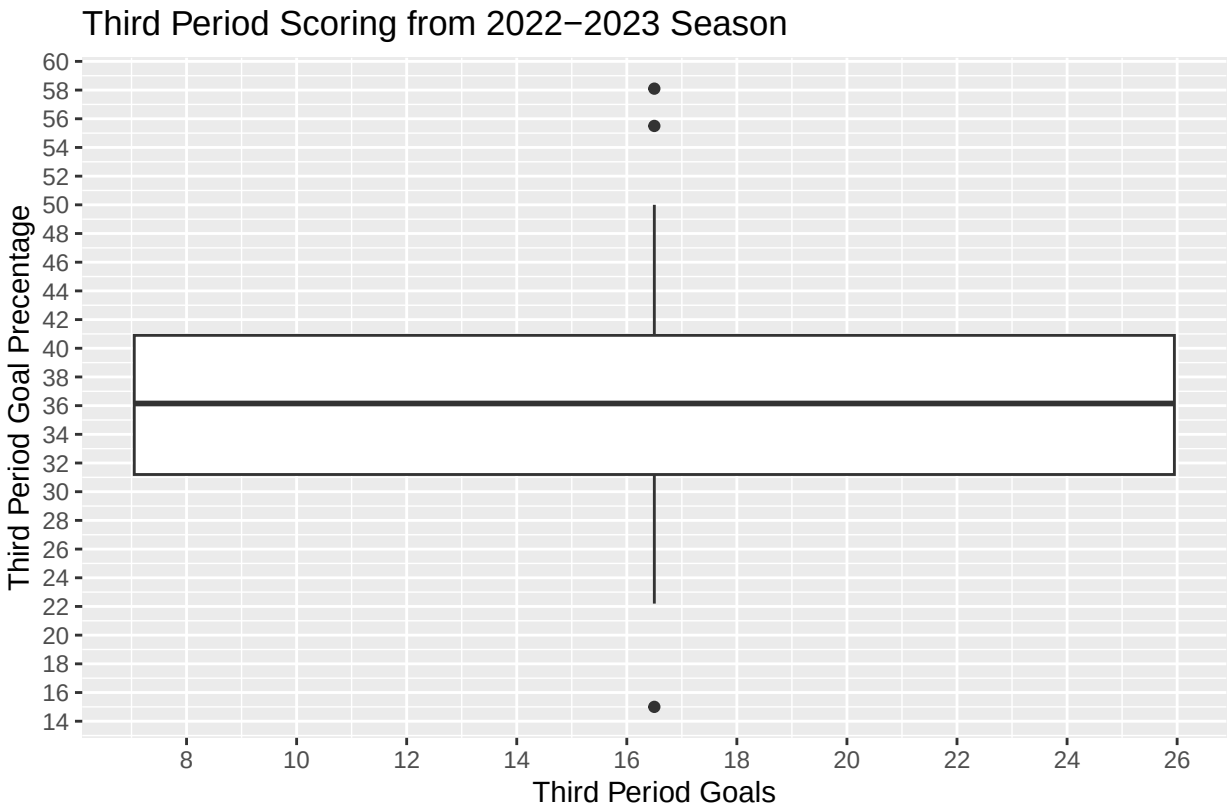
```
##
## Call:
## lm(formula = P2G ~ P2GPRCT, data = Goals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1386 -1.6471 -0.4596  0.6842  7.8478
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.60357    1.53459  -1.697   0.0963 .
## P2GPRCT      0.47294    0.04545  10.406 6.78e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.429 on 48 degrees of freedom
## Multiple R-squared:  0.6929, Adjusted R-squared:  0.6865
## F-statistic: 108.3 on 1 and 48 DF,  p-value: 6.776e-14
```

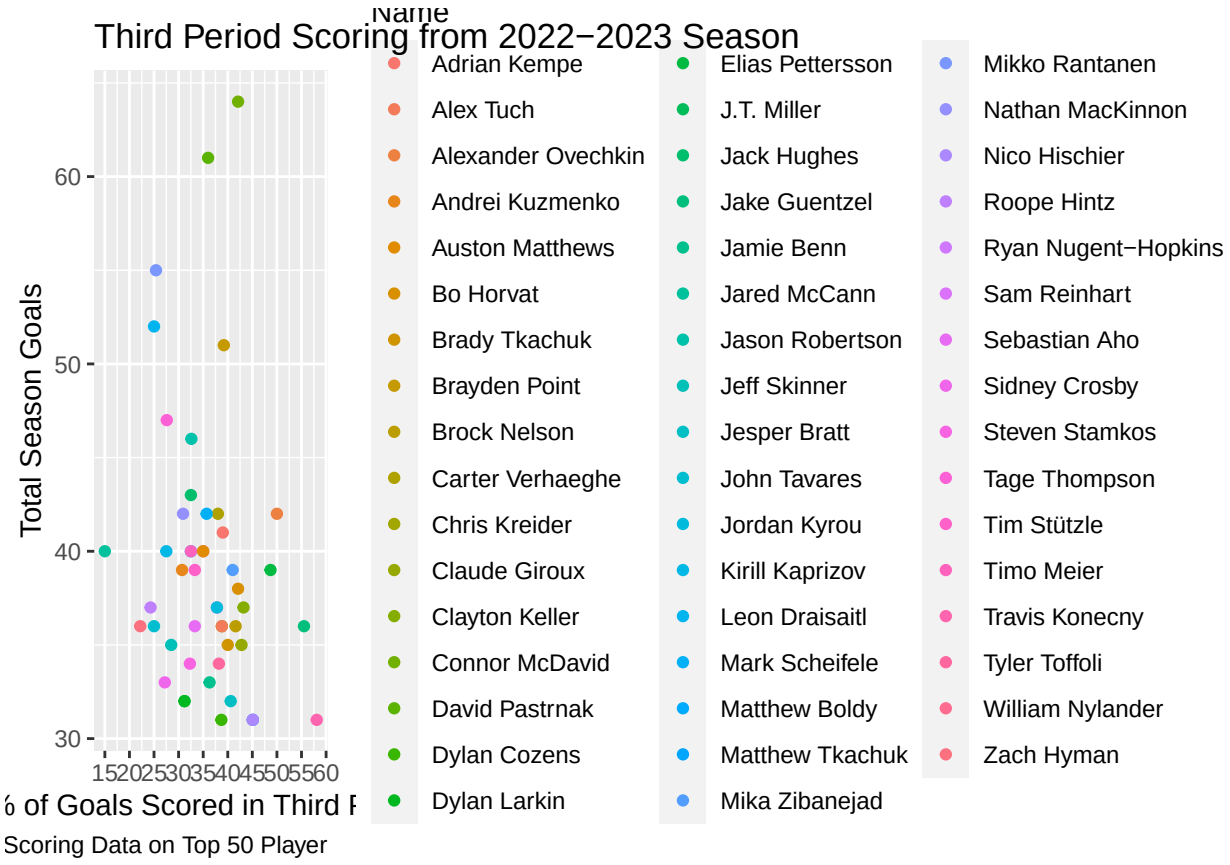

Residuals vs Fitted
Residuals
Fitted values
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance

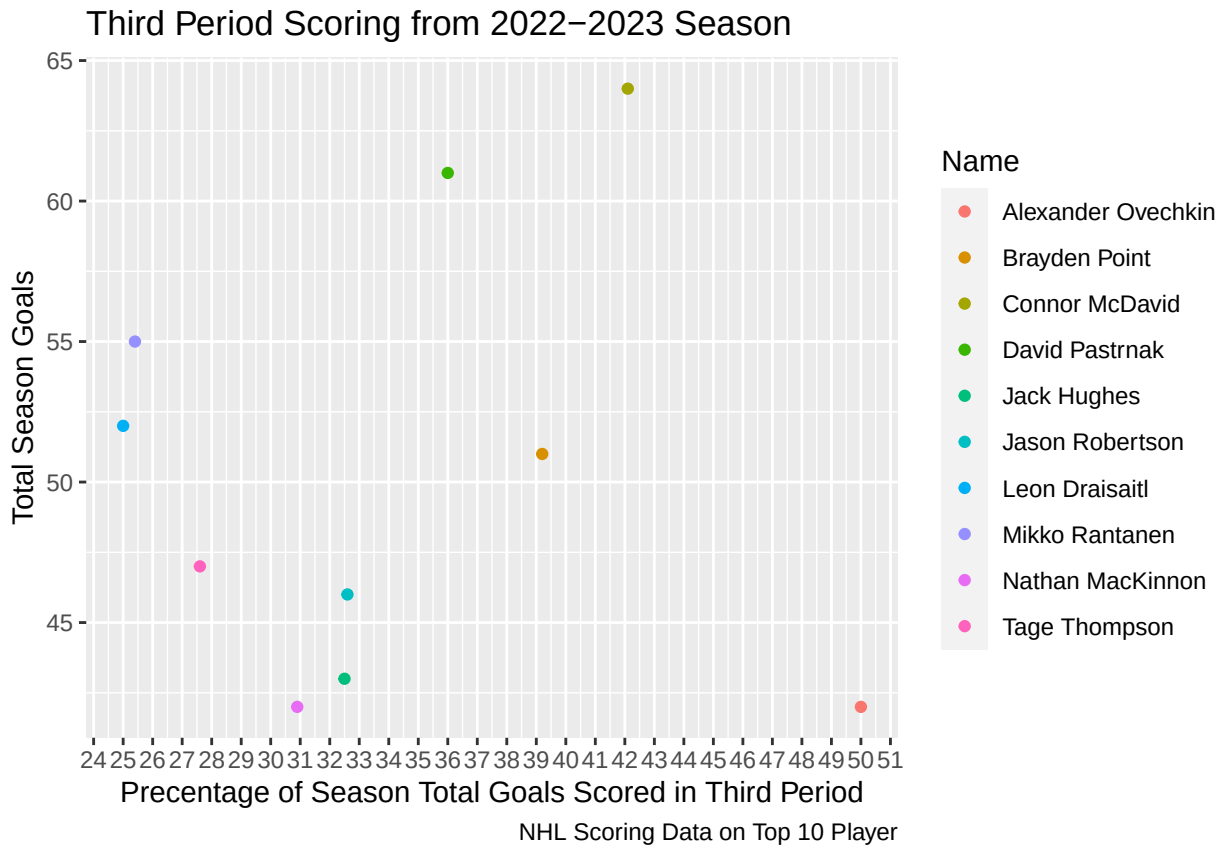
The third period data shows that the data following a linear projection with few outliers.
The third period Statistics showed that most players fell within the 1st(12.00) and 3rd(15.00) quartile, with only 3 outliers in the data set.
The Mean was 14.00 goals and the Median was 14.00 goals.
The one player that had the highest percentage of third period goals was Travis Konecny (58.1).
The player with the most third period goals was Connor McDavid, whom is a top 10 scorer (27).
The Top 10 Scores did not follow a linear projection in percentage, with a range from 25.4 to 50.0 percent.
All but four of the Top 10 Scores were below 33%, indicating the third period on average had more goals than the first period.




```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```

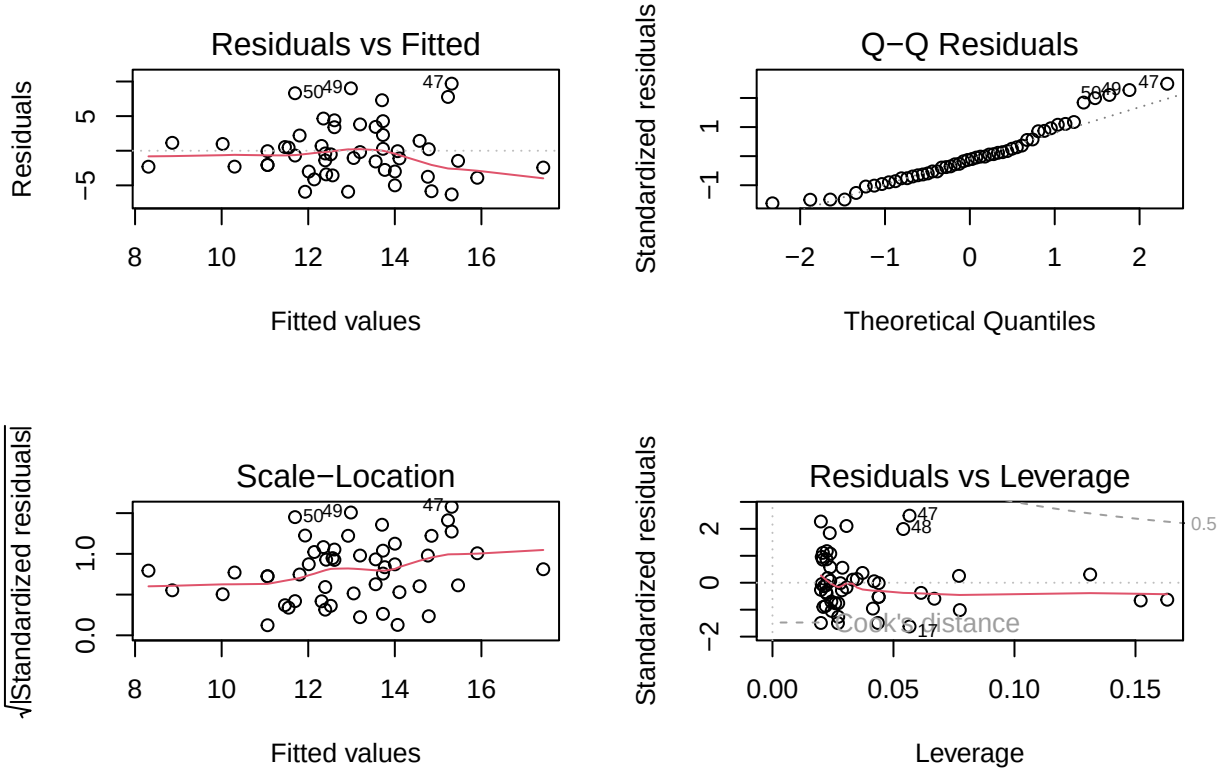

Third Period Scoring from 2022−2023 Season
Third Period Goal Precentage
60
58
56
54
52
50
48
46
44
42
40
38
36
34
32
30
28
26
24
22
20
18
16
14
8
10
12
14
16
18
20
22
24
26
Third Period Goals
NHL Scoring Data on Top 50 Player


Third Period Scoring from 2022−2023 Season
Name
Adrian Kempe
Alex Tuch
Alexander Ovechkin
Andrei Kuzmenko
Auston Matthews
Bo Horvat
Brady Tkachuk
Brayden Point
Brock Nelson
Carter Verhaeghe
Chris Kreider
Claude Giroux
Clayton Keller
Connor McDavid
David Pastrnak
Dylan Cozens
Dylan Larkin
Elias Pettersson
J.T. Miller
Jack Hughes
Jake Guentzel
Jamie Benn
Jared McCann
Jason Robertson
Jeff Skinner
Jesper Bratt
John Tavares
Jordan Kyrou
Kirill Kaprizov
Leon Draisaitl
Mark Scheifele
Matthew Boldy
Matthew Tkachuk
Mika Zibanejad
Mikko Rantanen
Nathan MacKinnon
Nico Hischier
Roope Hintz
Ryan Nugent−Hopkins
Sam Reinhart
Sebastian Aho
Sidney Crosby
Steven Stamkos
Tage Thompson
Tim Stützle
Timo Meier
Travis Konecny
Tyler Toffoli
William Nylander
Zach Hyman
Total Season Goals
60
50
40
30
15 20 25 30 35 40 45 50 55 60
6 of Goals Scored in Third P
Scoring Data on Top 50 Player




```
## [1] 0.7051229
```

```
##
## Call:
## lm(formula = P2G ~ P3GPRCT, data = Goals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3128 -2.6833 -0.4564  2.0065  9.6872
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.60233    2.55505   8.063 1.77e-10 ***
## P3GPRCT     -0.21158    0.06898  -3.067  0.00354 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.007 on 48 degrees of freedom
## Multiple R-squared:  0.1639, Adjusted R-squared:  0.1465
## F-statistic: 9.409 on 1 and 48 DF,  p-value: 0.003542
```


Residuals vs Fitted
Residuals
Fitted values
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
0.5

The summary data shows that a total of 1952 goals were scored by the top 50 scorers during the 2022-2023 season.
Of those goals, 544 goals were in the first period, 648 goal in the second period and 700 in the third period.
Further research supports the findings that, 22 of the players had the highest percentage of their goals in the third period, followed by 12 players in the second period and finally just 9 players in the first period.
The summary data for the first period shows out of the 544 total goals scored during that period, only 166 goals were scored by players with a scoring percentage higher than 35.5%
Next, the summary data for the second period shows out of the 648 total goals scored, 240 goals were scored by players with a scoring percentage higher than 35.5%
Finally, the summary data for the third period shows out of the 700 total goals scored, at total 362 goals were scored by players with a scoring percentage higher than 35.5%
Additionally, there were 7 players that had consistent scoring in each period with a percentage between 29.4 and 35.0.
This shows a clear correlation between the players having a tendency to score more goals in a specific period, with a total of 82% of players fitting in this pattern.
Of the top ten scorers, only Nathan MacKinnon was in the consistent scorer group.

List of Top Scorers per period:
1st Period Highest Percentage: Jared McCann 47.55
1st Period Most Goals: Tage Thompson 22
2nd Period Highest Percentage: Leon Draisaitl 48.0
2nd Period Most Goals: Leon Draisaitl 25
3rd Period Highest Percentage: Travis Konecny 58.1
3rd Period Most Goals: Connor McDavid 27

Data Sources:
https://www.quanthockey.com/nhl/seasons/2022-23-nhl-players-stats.html
https://www.hockey-reference.com/leagues/NHL_2023_games.html
https://github.com/ssaxton27/115_git_practice/

Project Summary:

I chose the data sets from quanthockey.com and hockey-reference.com because when combined the data set would be the most accurate and complete source for statistical information related to a specific hockey season. The combination of the two data sets covered about 85% of the information I need to complete the project. The primary reason I selected this project was the fact I have an interest in hockey and I have never seen a break down of scoring habits by period for hockey.

During the data cleaning and wrangling, the biggest huddle I encountered was having to go through every box score for all 82 games each player participated in. The process opened up the project to human error factors. This was countered and verified by using a simple algorithm to verify the sum of the period goals equaled the total goal count.

The "Big Question" for me is, do hockey players have a tendency to score more goals in a specific period or is scoring a random event. The data helped answer the question by clearly showing that players do tend to have a period with increased scoring.

The exploratory data provided some very informative data after some very simple cleaning steps. The patterns started to appear very early in the research. The data supported the findings that most players do have a period in which they score more goals than others. The data also highlighted that some players are at extremes in comparison to other players.

The methodology I chose was based on the data available and how it could be applied to the question I was asking. I looked at several different ways to process that data, but in the end, it was best analyzed by keeping the methodology simple and straight forward so as the results could be easily duplicated.

The final conclusion for this project was that the majority of the top 50 scorers in the National Hockey League do in fact have a period when they score most of their goals. The data shows that 82% of the players have a period in which they score more than 35.5% of their season total goals. One player in particular scored 58.1% of his goals in the 3rd period. Additionally, there were only 7 players that consistently scored goals in each period. Those players had a percentage between 29.4 and 35.0 in each period of the game.

In the future it would be interesting to revisit this project and dive deeper into the data set to correlate under what game conditions the goals were scored. The data set has the needed data, just formulating a algorithm and methodology to complete this additional work would have been outside the scope of this particular project.