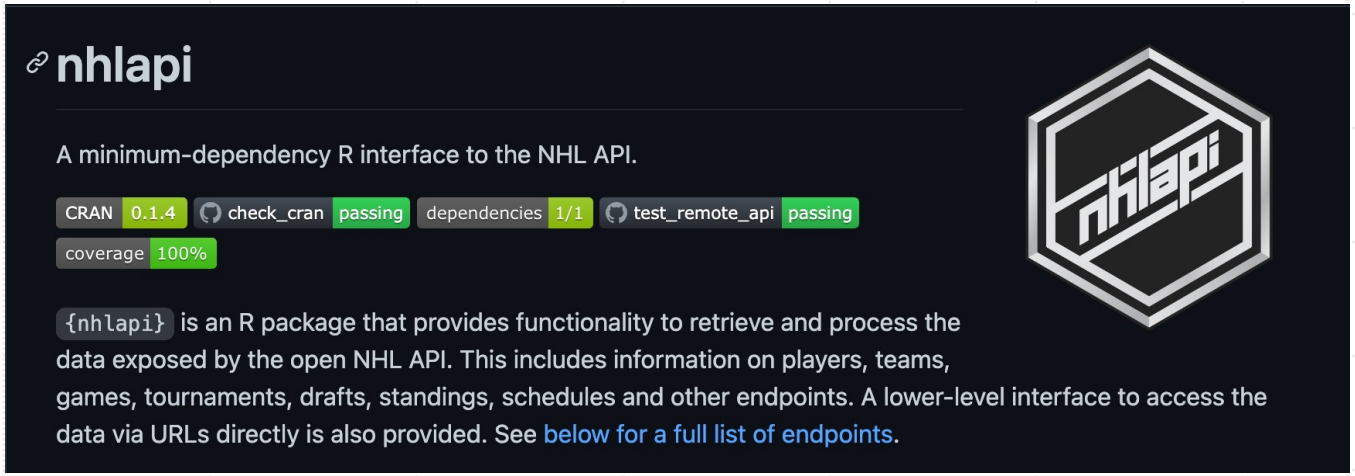
Examining Relationships in NHL Data Using nhlapi

By: Graydon Bower

nhlapi Package Information

- Used nhlapi package for R to retrieve data for 5 different NHL players between 2019 and 2021 NHL seasons: Patrice Bergeron, Patrick Kane, Steven Stamkos, Sidney Crosby & Connor McDavid
- API provided functionality to retrieve and process data exposed by open NHL AP
- Endpoint used for project was nhl_players_allseasons() (extracted info on player stats for each year of their careers based on player name/id from the NHL API)



- Links to where information used in project came from:
  - Bergeron: https://statsapi.web.nhl.com/api/v1/people/8470638/stats?stats=yearByYear
  - Kane: https://statsapi.web.nhl.com/api/v1/people/8474141/stats?stats=yearByYear
  - Stamkos: https://statsapi.web.nhl.com/api/v1/people/8474564/stats?stats=yearByYear
  - Crosby: https://statsapi.web.nhl.com/api/v1/people/8471675/stats?stats=yearByYear
  - McDavid: https://statsapi.web.nhl.com/api/v1/people/8478402/stats?stats=yearByYear

Research Questions Addressed

Which player has recorded the highest number of goals, assists and points during their careers?

Who had the highest point per game rate for each season from the 2019 season through the 2021 season?

For each player, how do they score the majority of their points? Does it come from goals or assists?

What was the distribution of penalty minutes like for these players between 2019 and 2021?

What is the relationship between penalty minutes and Plus/Minus Rate for each of these players?

Who recorded the highest Plus/Minus for each season from 2019 to 2021?

Who recorded the highest number of penalty minutes last season (2021)?

Prepping the Data for Use

- Created vector of player names called playerNames
- Created result1 object which joined nhl_players(playerNames) on nhl_players_allseasons(playerNames) endpoint on basis of “id” from nhl_players(playerNames) = “playerId” from nhl_players_allseasons(playerNames)
- Did the same thing for result1 through result 9
- Also filtered each result object by league.name == “National Hockey League” to obtain only NHL data
- Additional filter for seasonStart %in% c(2019, 2020, 2021) used in result4, result5, result6, result7 & result8 objects
- Additional filter for seasonStart %in% c(2021) used in result9

```
# Task 1: Creating line chart displaying the total number of cumulative points that Patrice Bergeron, Patrick Kane, Steve
# have accumulated over the courses of their careers so far.
playerNames <- c("Patrice Bergeron", "Patrick Kane", "Steven Stamkos", "Sidney Crosby", "Connor McDavid") # Creating a ve
library(dplyr)
library(highcharter)
result1 <- nhl_players(playerNames) %>%
  left_join(
    nhl_players_allseasons(playerNames),
    by = c("id" = "playerId") # Joining the data from the nhl_players() function provided by nhlapi on the nhl_players_al
  ) %>%
  filter(league.name == "National Hockey League") %>% # Filtering player data for only data from the NHL and not any othe
```

```
result4<- nhl_players(playerNames) %>%
  left_join(
    nhl_players_allseasons(playerNames),
    by = c("id" = "playerId") # Joining the data from the nhl_players() function provided by nhlapi on the nhl_players_al
  ) %>%
  filter(league.name == "National Hockey League", seasonStart %in% c(2019, 2020, 2021)) %>% # Process is much the same as
```

```
# Task 9: Figuring out who has recorded the highest number of penalty minutes la
result9<- nhl_players(playerNames) %>%
  left_join(
    nhl_players_allseasons(playerNames),
    by = c("id" = "playerId") # Joining the data from the nhl_players() function
  ) %>%
  filter(league.name == "National Hockey League", seasonStart %in% (2021)) %>%
```

Prepping the Data for Use Cont.

- Data then selected to be displayed for each result object using select() (fullName and seasonStart were selected for each object)
- Stat or stats included in select() differed from object to object, depending on what was being examined
- Data then saved to a tibble grouped by player full name
- Some data mutated to create new data points (specifically for result4 and result5)
- Finally, data is pipped into various charts/graphs to provide visualizations of data results

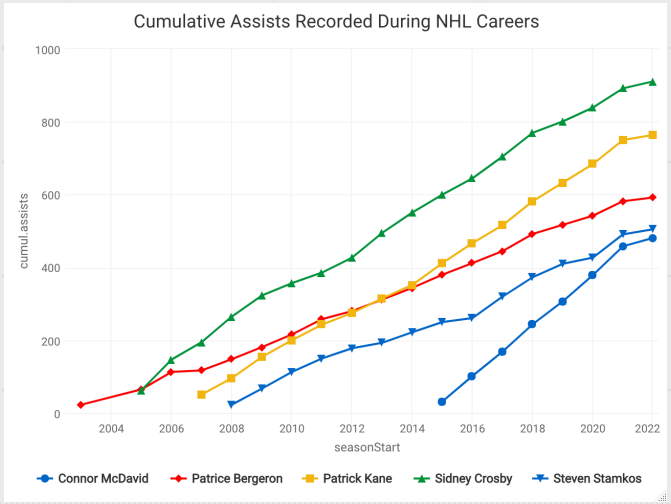
```r
# Task 4: Creating a column chart displaying the point per game rate for each player for each season from 2019 until 202
result4<- nhl_players(playerNames) %>%
  left_join(
    nhl_players_allseasons(playerNames),
    by = c("id" = "playerId") # Joining the data from the nhl_players() function provided by nhlapi on the nhl_players_al
  ) %>%
  filter(league.name == "National Hockey League", seasonStart %in% c(2019, 2020, 2021)) %>% # Process is much the same as
  select(fullName, seasonStart, stat.points) %>%
  as_tibble() %>%
  group_by(fullName) %>%
  mutate(pts.per.game = (stat.points/82)) %>% # Additionally, I created the pts.per.game stat by dividing the amount of p
  ungroup()
result4

result4 %>%
  hchart("column", hcaes(seasonStart, pts.per.game, group = fullName)) %>%
  hc_title(
    text = "Points Per Game Rate from 2019 to 2021") %>%
  hc_add_theme(hc_theme_google()) # Column chart that shows the pts.per.game for each player during 2019
```

```r
# Task 5: Examining the relationship between assist per game ratio and goals per game ratio for each player from the 201
library(googleVis)
result5<- nhl_players(playerNames) %>%
  left_join(
    nhl_players_allseasons(playerNames),
    by = c("id" = "playerId") # Joining the data from the nhl_players() function provided by nhlapi on the nhl_players_al
  ) %>%
  filter(league.name == "National Hockey League", seasonStart %in% c(2019, 2020, 2021)) %>%
  select(fullName, seasonStart, stat.goals, stat.points, stat.assists) %>%
  as_tibble() %>%
  group_by(fullName) %>%
  mutate(goals.per.game = (stat.goals/82)) %>%
  mutate(seasonStart.char = as.character(seasonStart)) %>%
  mutate(points.per.game = (stat.points/82)) %>%
  mutate(assists.per.game = (stat.assists/82)) %>%
  ungroup()
result5
```

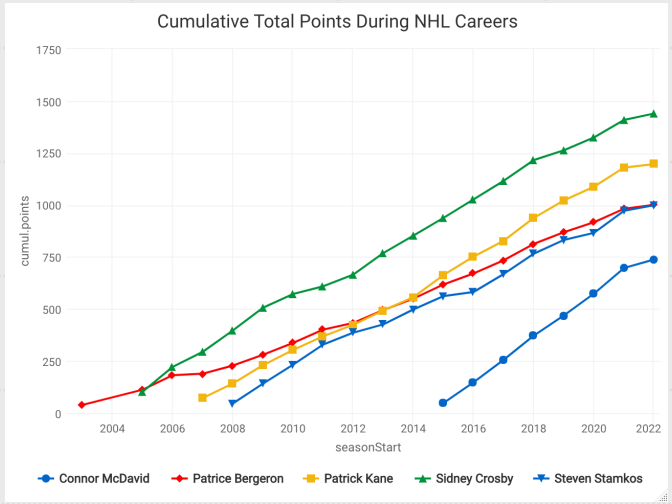
The rest of the result objects (Result1, Result2, Result3, Result5, Result6, Result7, Result8 & Result9) can be found in the project code.

Line Charts Produced Examining Cumulative Assists, Points and Goals

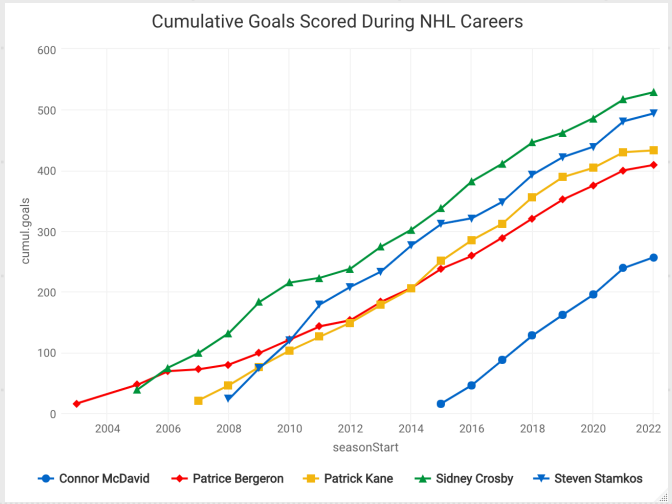
- Crosby has highest totals in assists, goals & points during his career out of 5 players examined
- McDavid appears to have the highest assist rate has only been in the NHL 7 years and is already about to surpass Stamkos in terms of career assists
- McDavid has been in the NHL for a smaller amount of time compared to the rest of these players
- Stamkos more of a goal scorer than playmaker
- Bergeron seems to have had most steady career (no real fluctuations in line charts)
- Not huge separation in terms of goal-scoring between each player(except McDavid) but Crosby and Kane are above the pack in terms of assists



Produced from result3 object



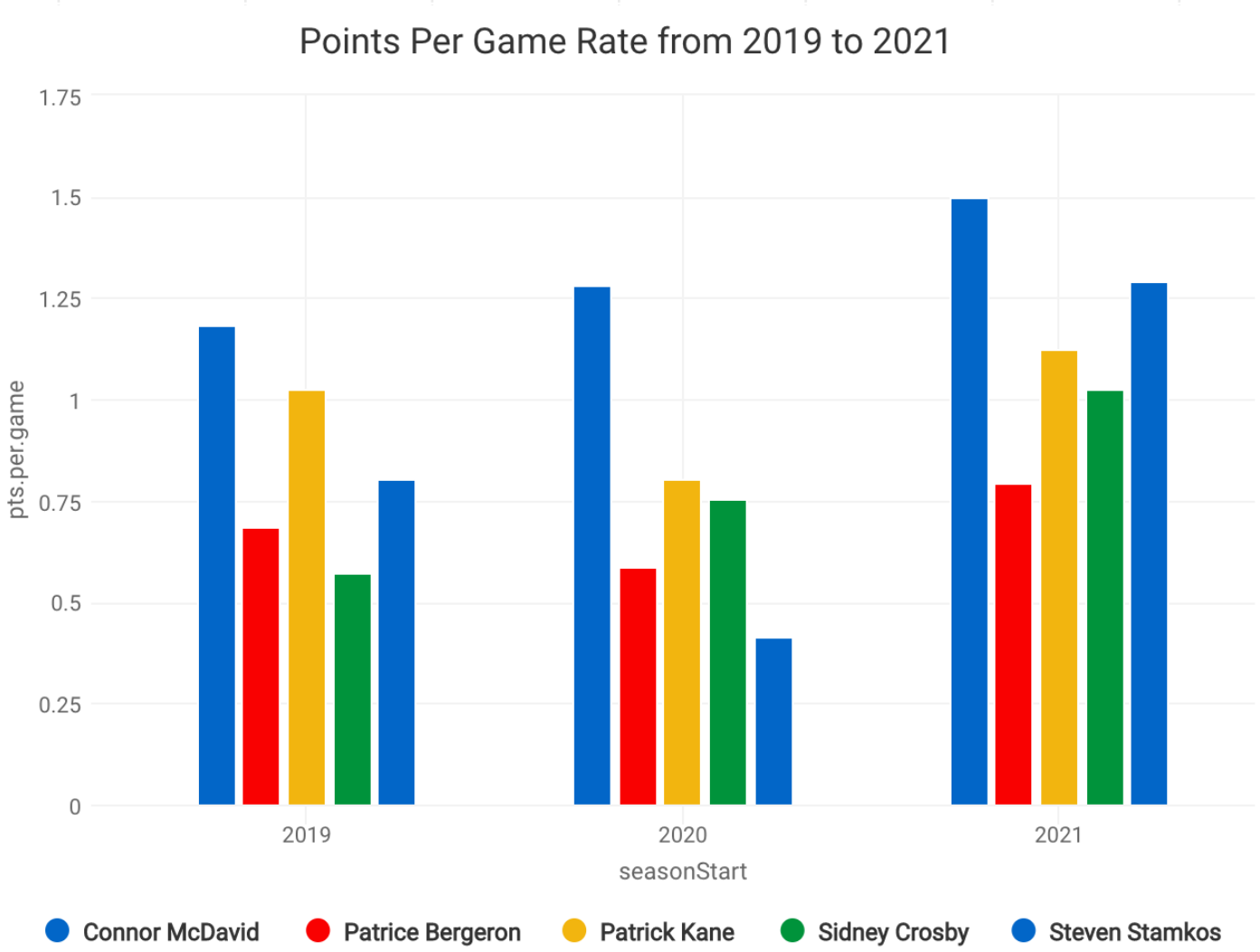
Produced from result1 object



Produced from result2 object

Column Chart Examining Point Per Game Rate

- McDavid has highest points per game rate each season
- While points per game fluctuates for each other player, McDavid's is consistently above 1 (very impressive)
- Only four other instances between 2019 & 2021 when this occured (Kane in 2019 & 2021, Crosby in 2021, and Stamkos in 2021)
- Bergeron's point per game rate is the most consistent (no huge fluctuations in column height)
- Stamkos's points per game has the most variation (huge fluctuations in column height)
- Points seemed to come easily for each player in 2021 (highest point per game ratios recorded during this year)

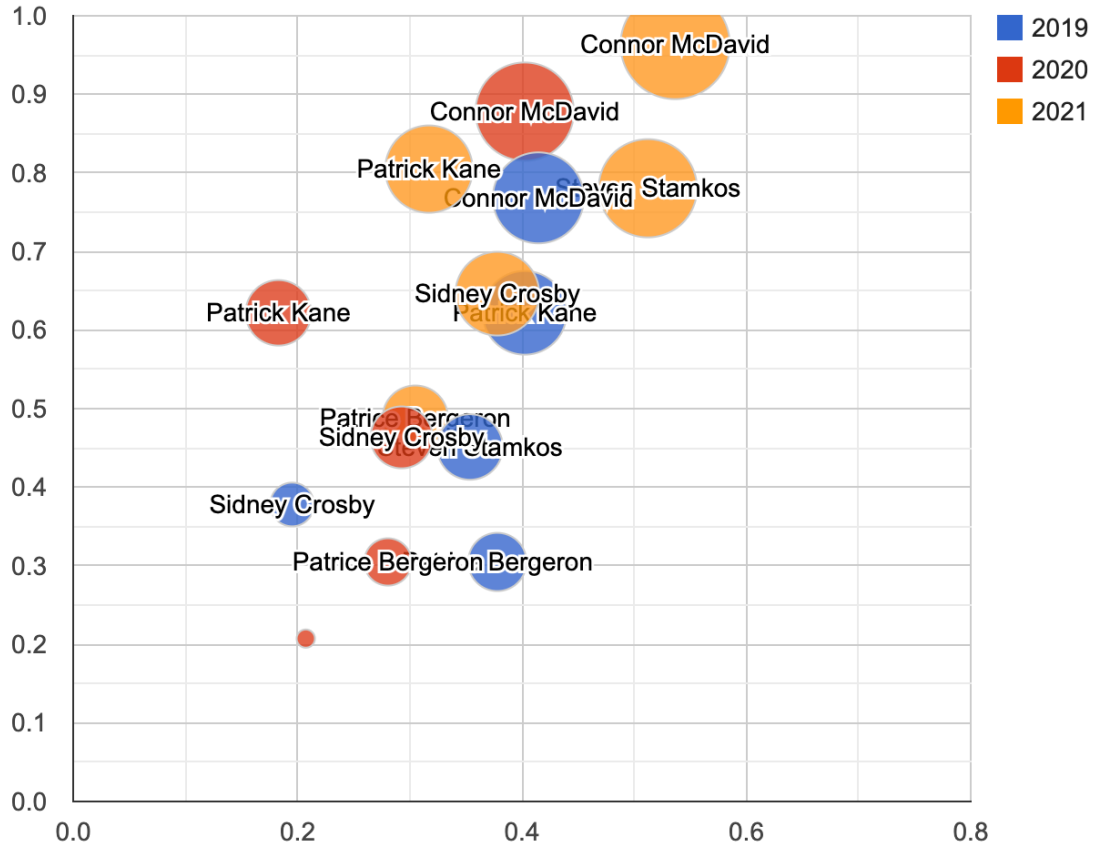


To clarify, 1st blue bar = McDavid, 2nd = Stamkos

Column chart produced from data in result4 object

Bubble Chart Examining Goals per Game vs. Assists per Game

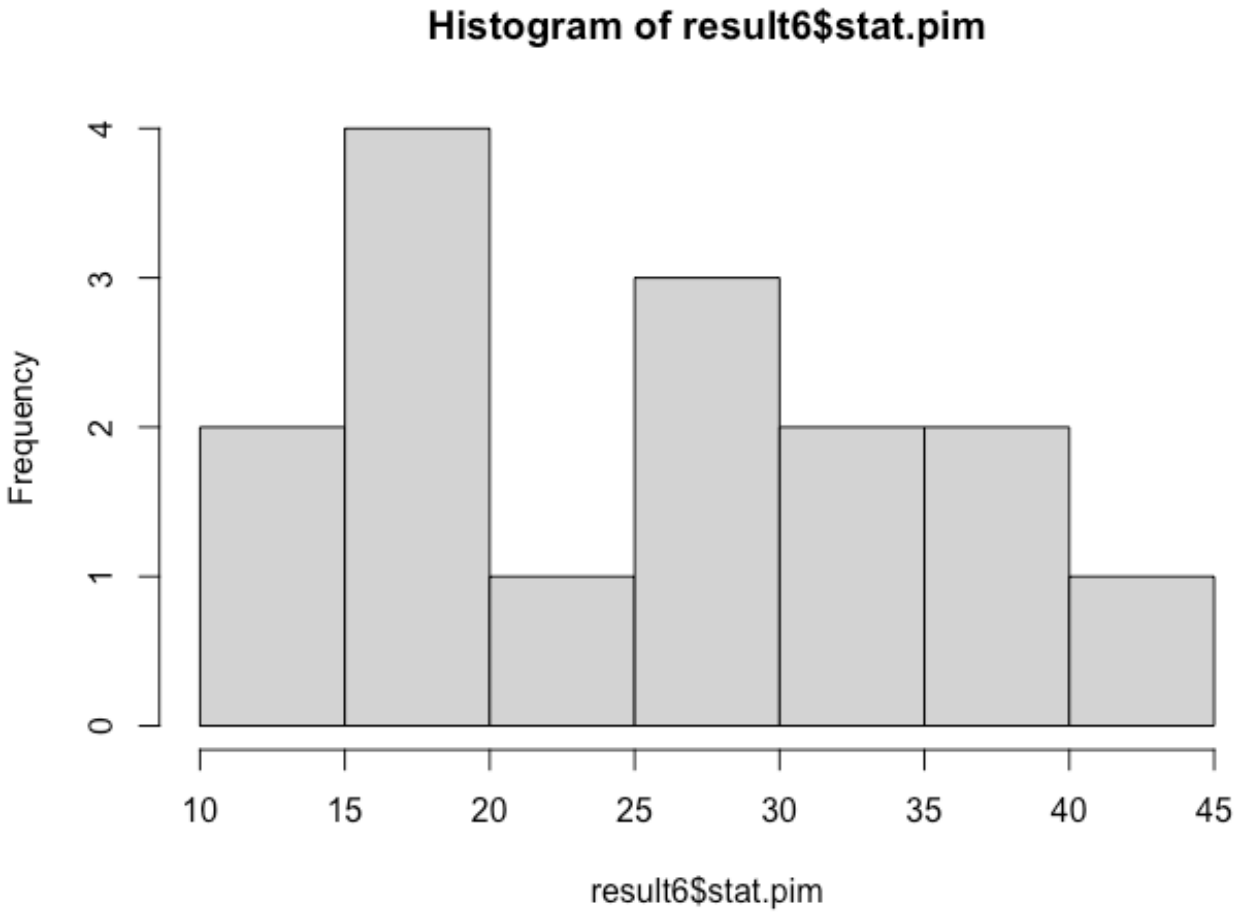
- Connor McDavid owns both the highest assist per game and goal per game ratios on the chart with his 2021 season (also highest points per game)
- He also owns 2 of the other 4 top point per game seasons
- Assists seem to equal goals in terms of which factor makes up the majority of Bergeron's points per game
- Kane also relies more heavily on assists to record points
- The same can be said for Stamkos & Crosby, although Stamkos recorded the 2nd highest goals per game ratio during his 2021 season
- Assists seem to be the main factor in accruing points in the NHL (at least for these 5 players)



Goals per game displayed on x-axis
Assists per game displayed on y-axis
Size of bubble = point per game ratio
Chart produced from data contained in result5 object

Histogram Examining Distribution of Penalty Minutes

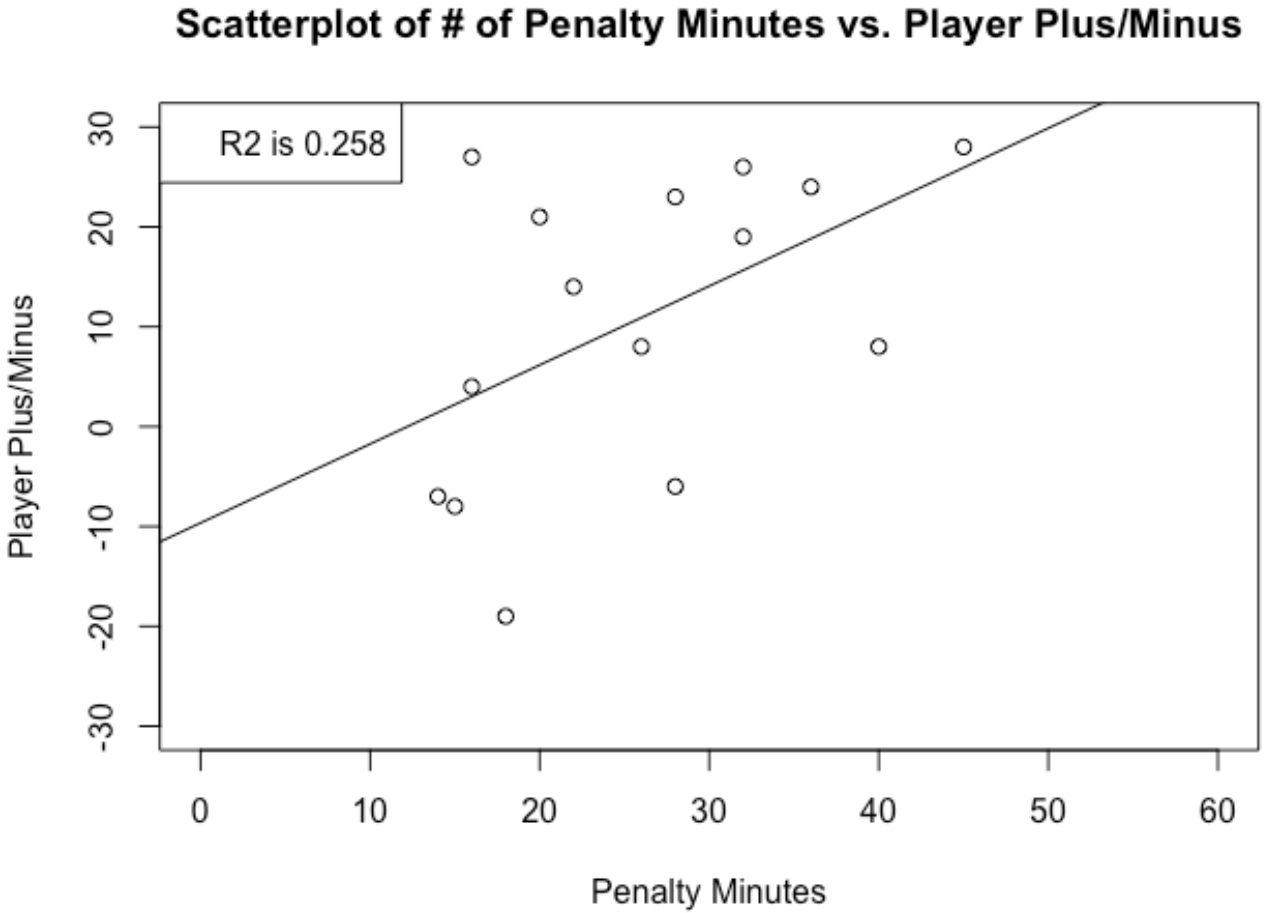
- Between 2019 and 2021, these players most frequently recorded between 15 and 20 penalty minutes per season
- The range of the data is between 10 and 45 penalty minutes
- The least frequent number of penalty minutes = tie (between 20 and 25 and between 40 and 45)
- This dataset appears to be relatively normally distributed
- The median of the dataset would be between 25 and 30 penalty minutes



Data used was taken for each player from 2019 season to 2021 season (each player shows up 3 times)
Histogram contains data for only the 5 players examined throughout project
Histogram created from data contained in result6 object

Scatterplot Examining Relationship Between # of Penalty Minutes and Player Plus/Minus Rating

- Created linear model to examine relationship between these two variables
- Used linear model to calculate R2 value; included in top left of scatterplot
- Used linear model to line of best fit for data using abline()
- Correlation is not very strong (weak positive relationship)
- Therefore, the number of penalty minutes doesn't have huge bearing on players Plus/Minus Ratio (at least for these 5 players)

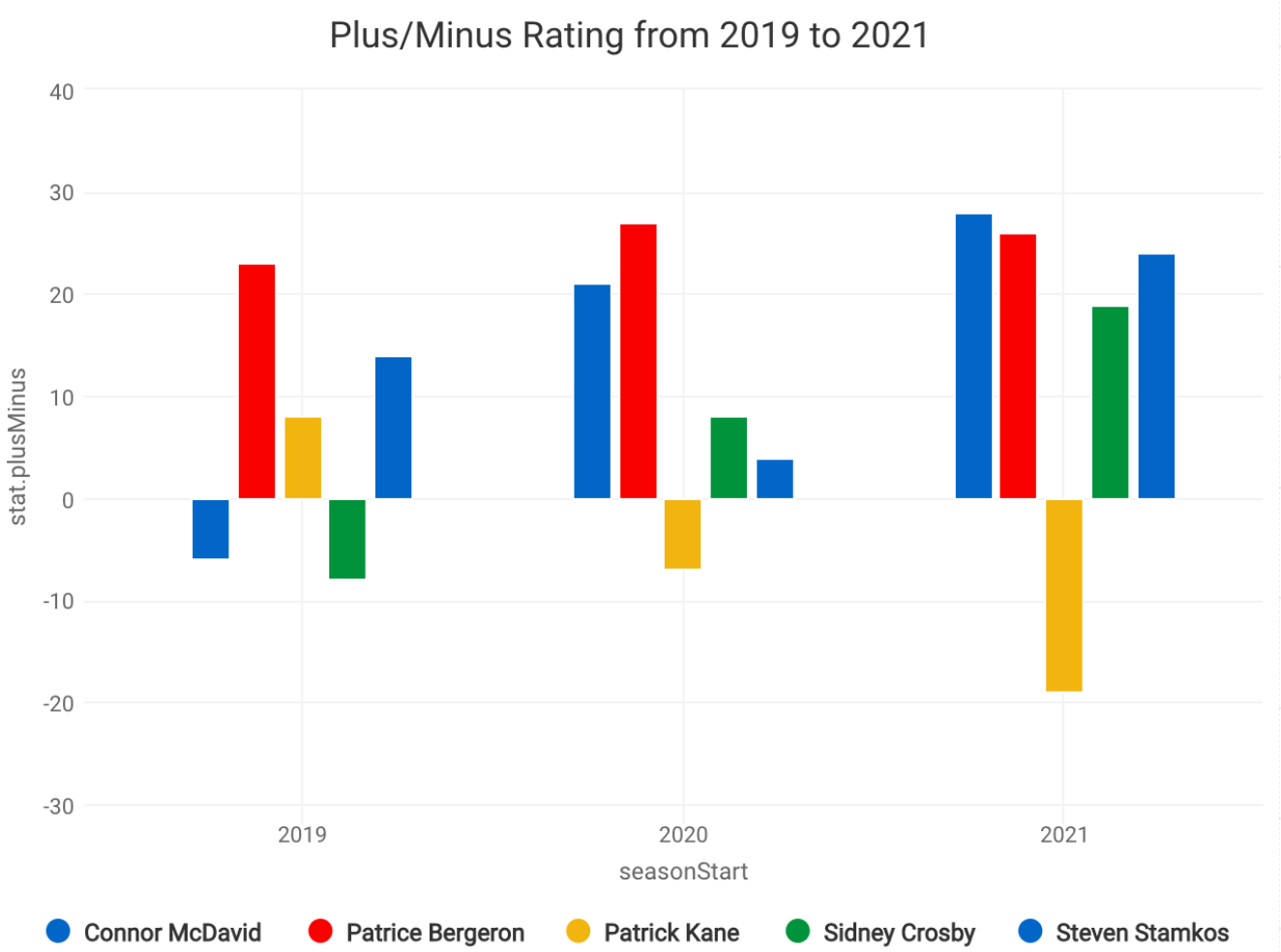


Again, data used was for each player from 2019 through 2021 (each player shows up 3 times in scatterplot)

Scatterplot created from data contained in result7 object

Column Chart Displaying Plus/Minus for 2019 through 2021

- There are four instances of seasons were negative plus/minuses were recorded (McDavid 2019, Crosby 2019, Kane 2020 & Kane 2021)
- Highest plus value on chart for both 2019 and 2020 = Patrice Bergeron (often has positive contribution on ice)
- McDavid has the highest plus/minus value on the chart (just below 30)
- Crosby's plus/minus has increasingly gotten better over these 3 years
- Kane's has progressively gotten worse (probably a symptom of the sub-par team he's on)
- Stamkos's plus/minus seems to fluctuate (not consistent at all)

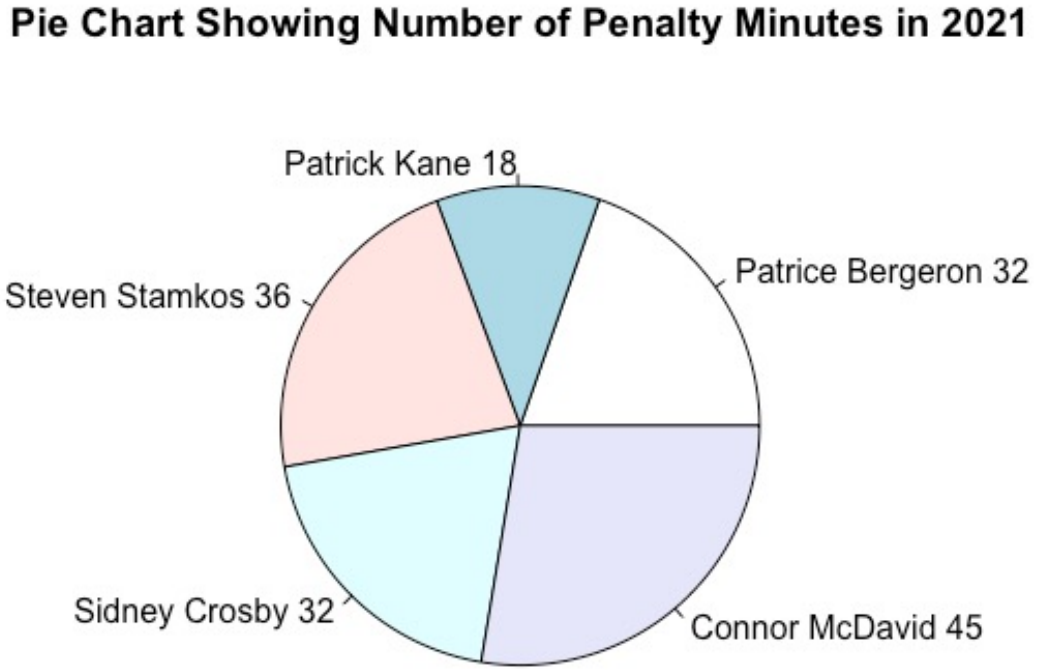


Again, 1st blue bar = McDavid, 2nd = Stamkos

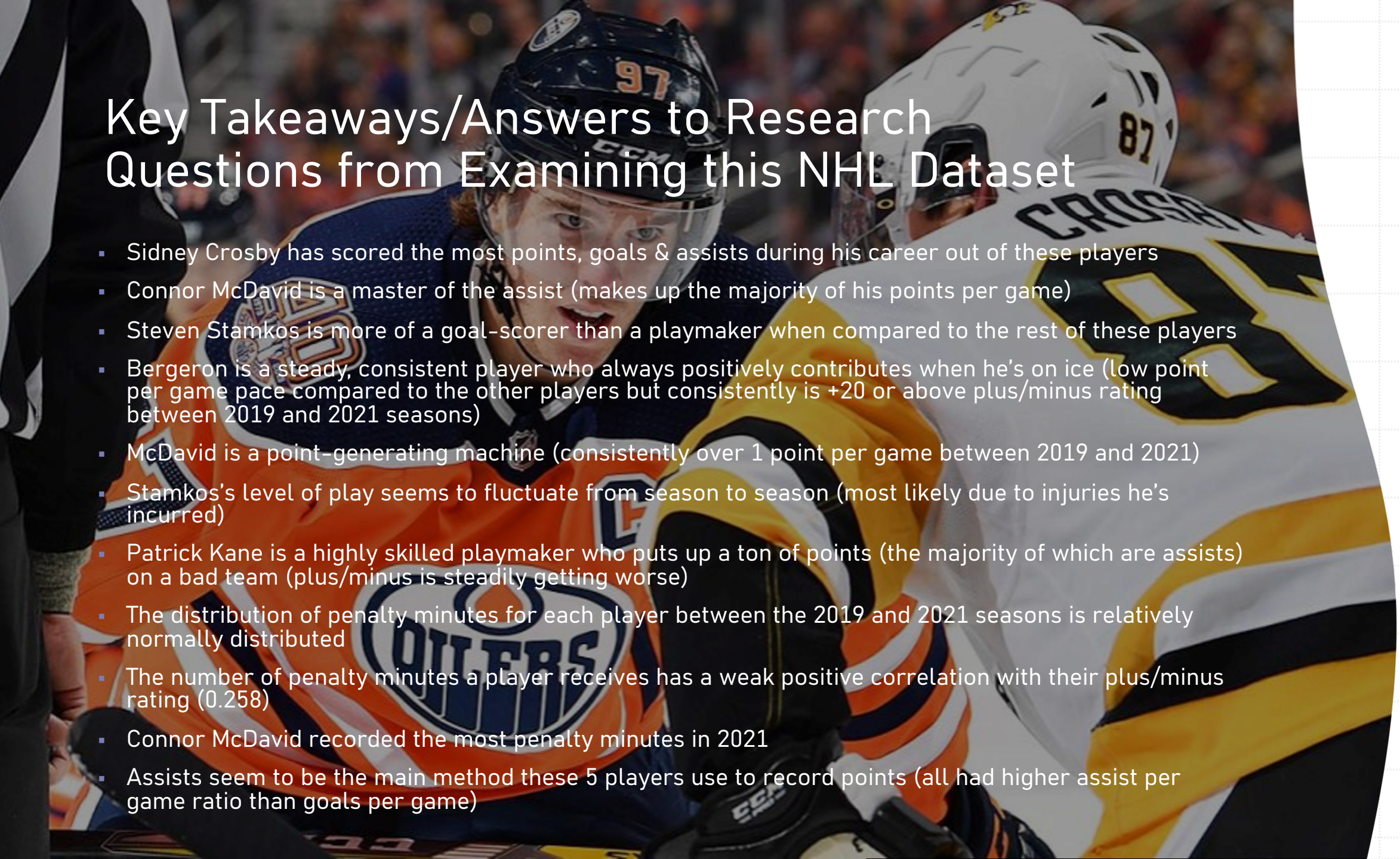
Column chart created from data contained in result8 object

Pie Chart Showing Penalty Minutes Recorded During 2021

- Connor McDavid recorded the most penalty minutes last season (45)
- Stamkos recorded the 2nd most (36)
- Crosby and Bergeron both recorded 32 penalty minutes
- Kane recorded the least amount of penalty minutes (18)



Pie chart created using data contained in result9 object



Key Takeaways/Answers to Research Questions from Examining this NHL Dataset

- Sidney Crosby has scored the most points, goals & assists during his career out of these players
- Connor McDavid is a master of the assist (makes up the majority of his points per game)
- Steven Stamkos is more of a goal-scorer than a playmaker when compared to the rest of these players
- Bergeron is a steady, consistent player who always positively contributes when he's on ice (low point per game pace compared to the other players but consistently is +20 or above plus/minus rating between 2019 and 2021 seasons)
- McDavid is a point-generating machine (consistently over 1 point per game between 2019 and 2021)
- Stamkos's level of play seems to fluctuate from season to season (most likely due to injuries he's incurred)
- Patrick Kane is a highly skilled playmaker who puts up a ton of points (the majority of which are assists) on a bad team (plus/minus is steadily getting worse)
- The distribution of penalty minutes for each player between the 2019 and 2021 seasons is relatively normally distributed
- The number of penalty minutes a player receives has a weak positive correlation with their plus/minus rating (0.258)
- Connor McDavid recorded the most penalty minutes in 2021
- Assists seem to be the main method these 5 players use to record points (all had higher assist per game ratio than goals per game)

Thank You!