Group project step 4

Craig Hatfield, Mike Natola, Noah Foilb, Brandon Gillis, Angantyr Gautason

11/24/2020

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(readxl)
library(readr)
library(ggplot2)
library(tidyr)
```

Load the files into Program. MAKE SURE THE EXCEL FILES ARE IN THE SAME FOLDER AS GROUP PROJECT.RMD

```
scoring_data_file <- "Scoring.csv"
scoring_data <- read.csv(scoring_data_file)

coaches_data_file <- "Coaches.csv"
coaches_data <- read.csv(coaches_data_file)

master_data_file <- "Master.csv"
master_data <- read.csv(master_data_file)

goalies_data_file <- "Goalies.csv"
goalies_data <- read.csv(goalies_data_file)
goalies_data[is.na(goalies_data)] <- 0

awards_players_data_file <- "AwardsPlayers.csv"
awards_players_data <- read.csv(awards_players_data_file)
```

Make a file to match players to their playerIDs

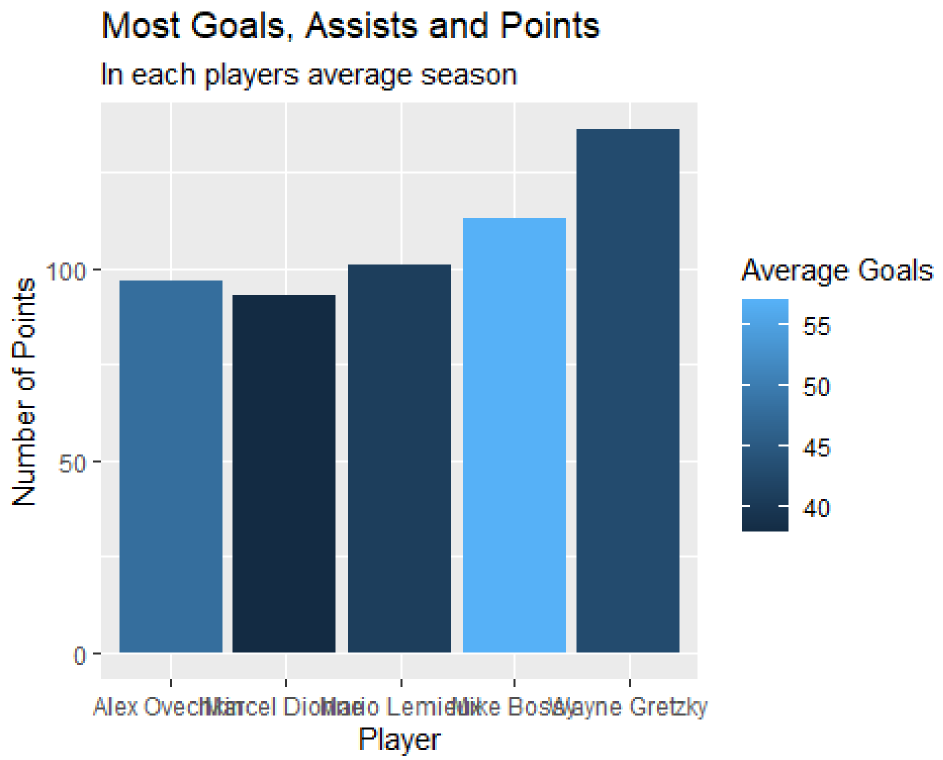
```
                    # Data set with PlayerIDs and names

Players <- master_data %>%                      # Assign master data to new variable
 select(playerID,                            # Only keep three columns from master data
      firstName,                           # These columns are playerID, firstName, lastName
      lastName)

                    # Data set with CoachIDs and names

Coaches <- master_data %>%                      # Assign master data to new variable
 select(coachID,                             # Only keep three columns from master data
      firstName,                           # These columns are playerID, firstName, lastName
      lastName)
```

#1.What player has the most goals,assists and points in thier average season? (Craig)

```
                    # Finding the Results

d1ID <- scoring_data %>%                        # Assign scoring data to new variable
 filter(lgID=="NHL")%>%                         # Filter data by players who are in the NHL
 group_by(playerID) %>%                         # Group the data by their player ID
 summarise(.groups = "drop",                    # Fix the ungrouping output error
 Average_Goal = round(sum(G/n())),                   # Average goals will be the summation of
their goals divided by the # of seasons
 Average_Assists = round(sum(A/n())),                # Same as ^ but with assists instead
 Average_Points = round(sum(Pts/n())),) %>%               # Same as ^ but with points instead
(Points is the goals + assists)
 arrange(desc(Average_Points)) %>%                   # Arrange by the most average points
 na.omit(d1ID)                               # Omit all Na's in dataset

                    # Displaying Names Instead of PlayerID

d1 <- left_join(d1ID,Players,"playerID") %>%          # Join together d1ID with Players
dataset to replace PlayerID with their names
 mutate(Name = paste(firstName,lastName)) %>%          # Join together the first and last
name in the Players
 select("Name" = Name,                          # Keep four variables
      "Average Goals" = Average_Goal,
      "Average Assists" = Average_Assists,
      "Average Points" = Average_Points)

                    # Displaying Results
head(d1,20)
```

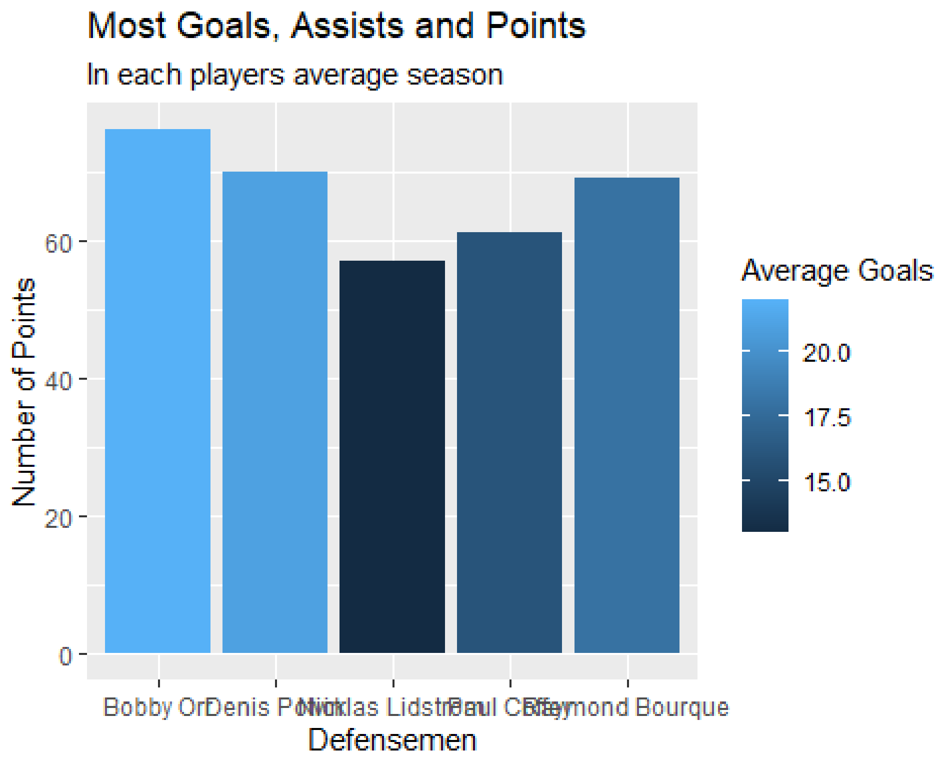
```
## # A tibble: 20 x 4
##    Name              `Average Goals` `Average Assists` `Average Points`
##    <chr>                       <dbl>             <dbl>            <dbl>
##  1 Wayne Gretzky                  43                93              136
##  2 Mike Bossy                     57                55              113
##  3 Mario Lemieux                  41                61              101
##  4 Alex Ovechkin                  48                49               97
##  5 Marcel Dionne                  38                55               93
##  6 Evgeni Malkin                  35                53               88
##  7 Sidney Crosby                  32                55               87
##  8 Jaromir Jagr                   35                52               87
##  9 Phil Esposito                  38                46               84
## 10 Dale Hawerchuk                 30                52               83
## 11 Joe Sakic                      31                51               82
## 12 Steven Stamkos                 45                38               82
## 13 Bobby Clarke                   24                57               81
## 14 Bernie Federko                 26                54               81
## 15 Guy Lafleur                    33                47               80
## 16 Steve Yzerman                  31                48               80
## 17 Bryan Trottier                 29                50               79
## 18 Jari Kurri                     33                44               78
## 19 Gilbert Perreault              30                48               78
## 20 Peter Stastny                  28                49               77
```

```
                  # Making Graph

ggplot(data = d1[1:5,],                          # Use the top five people from the d1 dataset
 aes(x = Name,                                   # X axis is for the names
 y = `Average Points`,                           # Y axis is for average points
 fill = `Average Goals`)) +                      # Fill color with average goals
 geom_bar(stat = "identity",
 position= "dodge")  +
 labs(title = "Most Goals, Assists and Points",          # Make title and subtitle
 subtitle = "In each players average season",
 x = "Player",                                   # Make x label and y label
 y = "Number of Points")
```



#2a.What defensemen scored the most points in their average season? (Craig)

```
                    # Finding the Results

d2aID <- scoring_data %>%                        # Assign scoring data to new variable
 filter(lgID=="NHL",                        # Filter data by players who are in the NHL
 pos=="D") %>%                              # And players who play defense
 group_by(playerID) %>%                          # Group by their playerIds
 summarise(.groups = "drop",                     # Fix the ungrouping output error
 G = round(sum(G/n())),                        # Average goals will be the summation of their
goals divided by the # of seasons
 A = round(sum(A/n())),                        # Same as ^ but with assists instead
 Pts = round(sum(Pts/n())),) %>%                     # Sum up their Points
 select(                                # Selects only the data we want to keep
 playerID,G,A,Pts) %>%                        # Only keep playerId, Goals, Assists, and Points
 arrange(desc(Pts))                          # Arrange by their points

                    # Displaying Names Instead of PlayerID

d2a <- left_join(d2aID,Players,"playerID") %>%          # Join together d2aID with Players
dataset to replace PlayerID with their names
 mutate(Name = paste(firstName,lastName)) %>%          # Join together the first and last
name in the Players
 select("Name" = Name,                         # Keep four variables
      "Average Goals" = G,
       "Average Assists" = A,
       "Average Points" = Pts)

                        # Displaying Results

head(d2a,20)

## # A tibble: 20 x 4
##    Name               `Average Goals` `Average Assists` `Average Points`
##    <chr>                        <dbl>             <dbl>            <dbl>
##  1 Bobby Orr                       22                54               76
##  2 Denis Potvin                    21                49               70
##  3 Raymond Bourque                 18                51               69
##  4 Paul Coffey                     16                45               61
##  5 Nicklas Lidstrom                13                44               57
##  6 Al MacInnis                     15                41               55
##  7 Phil Housley                    15                39               54
##  8 Brian Leetch                    13                41               54
##  9 Doug Wilson                     15                37               52
## 10 Pekka Rautakallio               11                40               51
## 11 Paul Reinhart                   12                39               51
## 12 Erik Karlsson                   12                37               50
## 13 Brad Park                       12                38               50
## 14 Larry Murphy                    11                37               49
## 15 Larry Robinson                  10                38               48
## 16 Sergei Zubov                    10                39               48
## 17 Brian Rafalski                   7                40               47
## 18 Gary Suter                      11                36               47
## 19 Borje Salming                    9                37               46
## 20 Mark Howe                       12                33               45

                        # Making Graph

ggplot(data = d2a[1:5,],                         # Use the top five people from the d2a dataset
 aes(x = Name,                                  # X axis is for the names
 y = `Average Points`,                           # Y axis is for average points
 fill = `Average Goals`)) +                      # Fill color with the average goals
 geom_bar(stat = "identity",
 position = "dodge")  +
 labs(title = "Most Goals, Assists and Points",         # Set title and subtitle
 subtitle = "In each players average season",
 x = "Defensemen",                                # Set x and y label
 y = "Number of Points")
```



#2b.What experienced defensemen has the best plus/minus? (Craig)

```
                    # Finding the Results

d2bID <- scoring_data %>%                     # Assign scoring data to new variable
 rename(plus_minus= "X...") %>%                    # Rename X... (Supposed to be +/-) to
plus_minus
 drop_na(plus_minus) %>%                      # Get rid of all the Na's in the data set
 filter(lgID=="NHL",                    # Filter by players in the NHL
 pos=="D") %>%                          # And by players who play defense
 group_by(playerID) %>%                       # Group data by their player IDs
 summarise(.groups = "drop",                   # Fix the ungrouping output error
 plus_minus=sum(plus_minus),                     # Plus_minus will be the summation of each
players +/- statistic
 GP=sum(GP))%>%                             # Amount of games played
 filter(GP>750) %>%                         # Experienced players will have at leasted played
750 games by our standards
 arrange(desc(plus_minus))                    # List by most plus_minus

                    # Displaying Names Instead of PlayerID

d2b <- left_join(d2bID,Players,"playerID") %>%         # Join together d2bID with Players
dataset to replace PlayerID with their names
 mutate(Name = paste(firstName,lastName)) %>%         # Join together the first and last
name in the Players
```

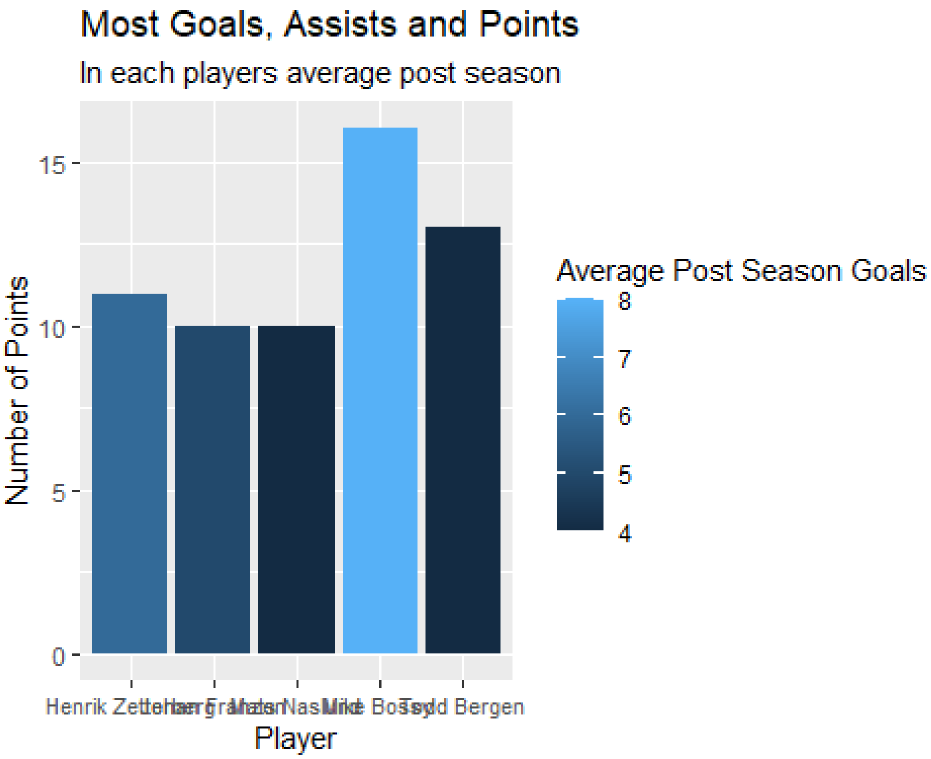
```
 select("Name" = Name,                         # Keep three variables
      "+/-" = plus_minus,
      "Games Played" = GP)

                    # Displaying Results

head(d2b,20)

## # A tibble: 20 x 3
##    Name           `+/-` `Games Played`
##    <chr>          <int>        <int>
##  1 Larry Robinson     730         1384
##  2 Raymond Bourque    528         1612
##  3 Denis Potvin       460         1060
##  4 Serge Savard       460         1038
##  5 Nicklas Lidstrom   450         1564
##  6 Brad McCrimmon     444         1222
##  7 Scott Stevens      393         1635
##  8 Mark Howe          390          866
##  9 Al MacInnis        373         1416
## 10 Brad Park          358         1113
## 11 Dallas Smith       355          773
## 12 Chris Chelios      350         1651
## 13 Guy Lapointe       329          884
## 14 Bill Hajt          321          854
## 15 Andre Dupont       299          800
## 16 Paul Coffey        294         1409
## 17 Rod Langway        277          994
## 18 Kevin Lowe         252         1254
## 19 Charlie Huddy      241         1017
## 20 Mike Ramsey        218         1070
```

#3. What player has the most goals,assists and points in thier average post-season?

```
                    # Finding the Results

d3ID <- scoring_data %>%                         # Assign scoring data to new variable
 filter(lgID=="NHL")%>%                          # Filter data by players who are in the NHL
 group_by(playerID) %>%                          # Group the data by their player ID
 summarise(.groups = "drop",                     # Fix the ungrouping output error
 Average_Goal = round(sum(PostG/n())),                # Average goals will be the summation of their goals divided by the # of seasons
 Average_Assists = round(sum(PostA/n())),             # Same as ^ but with assists instead
 Average_Points = round(sum(PostPts/n())),) %>%        # Same as ^ but with points instead (Points is the goals + assists)
 arrange(desc(Average_Points)) %>%                    # Arrange by the most average points
 na.omit(d1ID)                                   # Omit all Na's in dataset
```

```
                    # Displaying Names Instead of PlayerID

d3 <- left_join(d3ID,Players,"playerID") %>%            # Join together d3ID with Players
dataset to replace PlayerID with their names
  mutate(Name = paste(firstName,lastName)) %>%          # Join together the first and last
name in the Players
  select("Name" = Name,                             # Keep four variables
         "Average Post Season Goals" = Average_Goal,
         "Average Post Season Assists" = Average_Assists,
         "Average Post Season Points" = Average_Points)

                    # Displaying Results

head(d3,20)

## # A tibble: 20 x 4
##    Name        `Average Post Season~ `Average Post Season~ `Average Post Season~
##    <chr>                       <dbl>                 <dbl>                 <dbl>
##  1 Mike Bossy                      8                     8                    16
##  2 Todd Bergen                     4                     9                    13
##  3 Henrik Zet~                     6                     6                    11
##  4 Johan Fran~                     5                     5                    10
##  5 Mats Naslu~                     4                     6                    10
##  6 Pavel Dats~                     3                     6                     9
##  7 Nicklas Li~                     3                     6                     9
##  8 Hakan Loob                      4                     5                     9
##  9 Brian Rafa~                     3                     6                     9
## 10 Nicklas Ba~                     3                     5                     8
## 11 Michel Bri~                     5                     3                     8
## 12 Dickie Moo~                     3                     5                     8
## 13 Logan Cout~                     4                     3                     7
## 14 Bernie Geo~                     4                     4                     7
## 15 Chris Krei~                     5                     2                     7
## 16 Milan Lucic                     3                     4                     7
## 17 Larry Robi~                     1                     6                     7
## 18 Ryane Clowe                     3                     4                     6
## 19 Kjell Dahl~                     2                     4                     6
## 20 Gordie Dri~                     4                     2                     6

                    # Making Graph

ggplot(data = d3[1:5,],                          # Use the top five people from the d3 dataset
  aes(x = Name,                                  # X axis is for the names
  y = `Average Post Season Points`,               # Y axis is for average post season goals
  fill = `Average Post Season Goals`)) +            # Fill color with average post season goals
```

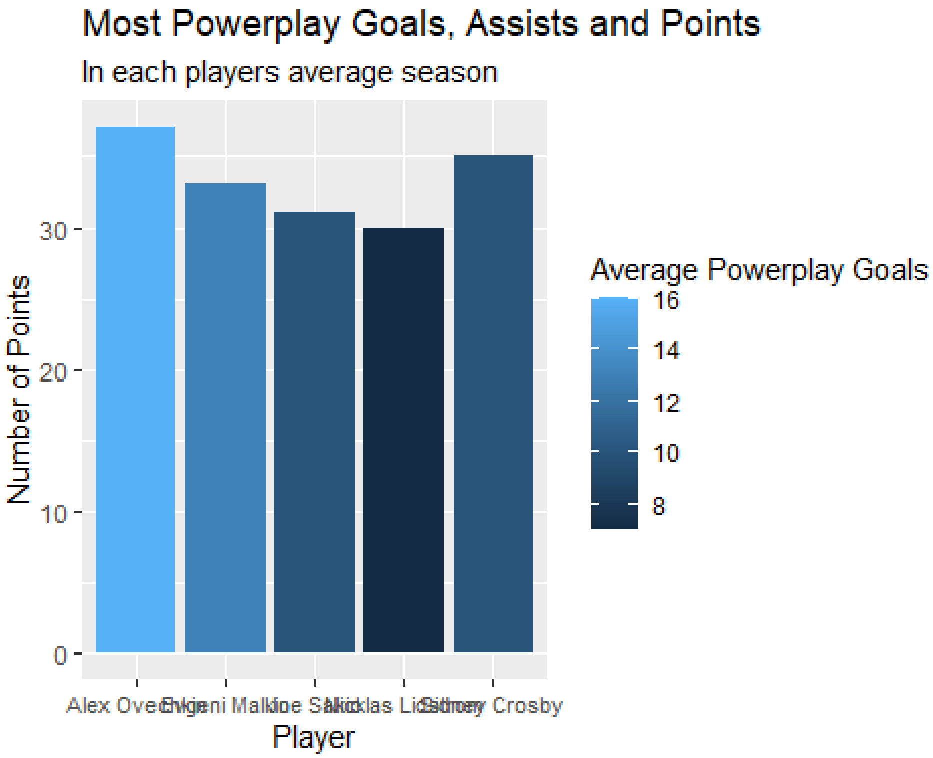
```
geom_bar(stat = "identity",
position= "dodge")  +
labs(title = "Most Goals, Assists and Points",          # Set title and subtitle
subtitle = "In each players average post season",
x = "Player",                                   # Set x label and y label
y = "Number of Points") +
theme(axis.text.x = element_text(size = 8))            # Text spacing for names
```



#4a. What player has the most goals,assists and points in thier average powerplay per season?

```
                    # Finding the Results

d4aID <- scoring_data %>%                         # Assign scoring data to new variable
 filter(lgID=="NHL")%>%                          # Filter data by players who are in the NHL
 group_by(playerID) %>%                          # Group the data by their player ID
 summarise(.groups = "drop",                      # Fix the ungrouping output error
 PPG = round(sum(PPG/n())),                        # Average goals will be the summation of their
goals divided by the # of seasons
 PPA = round(sum(PPA/n()))) %>%                          # Same as ^ but with assists instead
 mutate(PPP = PPA + PPG) %>%                              # New Column Called PPP which is PPA +
PPG
 arrange(desc(PPP)) %>%                            # Arrange by the most average points
 na.omit(d1ID)                               # Omit all Na's in dataset
```

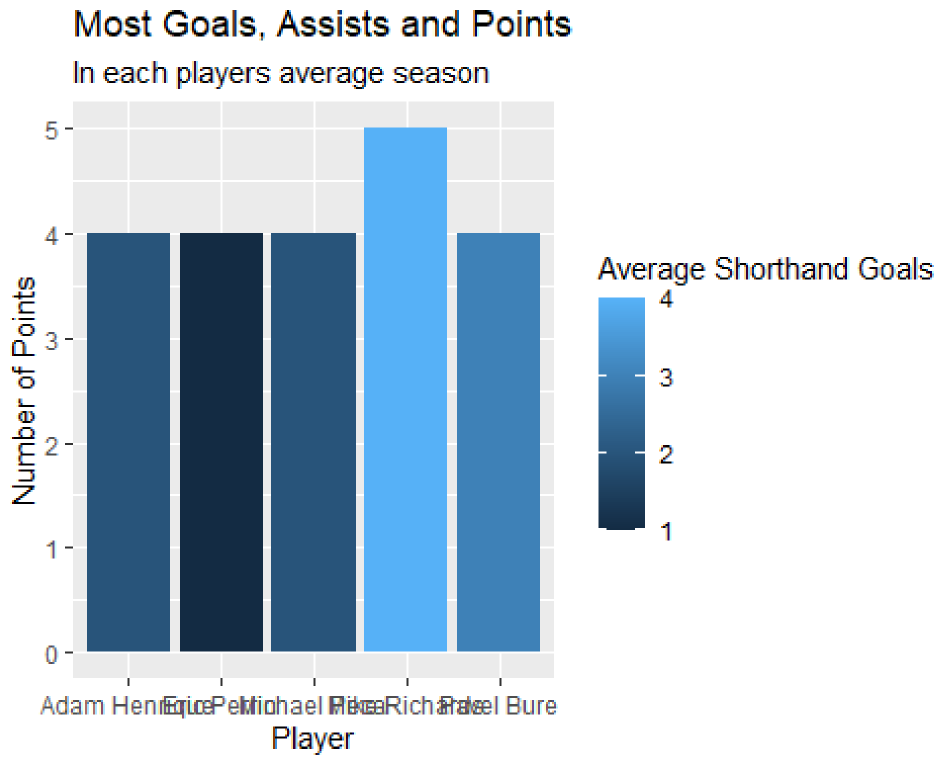
```
                    # Displaying Names Instead of PlayerID

d4a <- left_join(d4aID,Players,"playerID") %>%          # Join together d4aID with Players
dataset to replace PlayerID with their names
  mutate(Name = paste(firstName,lastName)) %>%          # Join together the first and last
name in the Players
  select("Name" = Name,                              # Keep four variables
        "Average Powerplay Goals" = PPG,
        "Average Powerplay Assists" = PPA,
        "Average Powerplay Points" = PPP)

                    # Displaying Results

head(d4a,20)

## # A tibble: 20 x 4
##    Name        `Average Powerplay G~ `Average Powerplay A~ `Average Powerplay P~
##    <chr>                     <dbl>                 <dbl>                 <dbl>
##  1 Alex Ovech~                  16                    21                    37
##  2 Sidney Cro~                  10                    25                    35
##  3 Evgeni Mal~                  13                    20                    33
##  4 Joe Sakic                    10                    21                    31
##  5 Nicklas Li~                   7                    23                    30
##  6 Steven Sta~                  16                    14                    30
##  7 Nicklas Ba~                   7                    22                    29
##  8 Ryan Getzl~                   8                    21                    29
##  9 Jaromir Ja~                  10                    19                    29
## 10 Brian Leet~                   6                    23                    29
## 11 Dany Heatl~                  14                    14                    28
## 12 Anze Kopit~                   9                    19                    28
## 13 Brad Richa~                   7                    21                    28
## 14 Ilya Koval~                  12                    15                    27
## 15 Teemu Sela~                  12                    15                    27
## 16 Pavel Dats~                   7                    19                    26
## 17 Patrick Ka~                   8                    18                    26
## 18 Paul Kariya                   9                    17                    26
## 19 Eric Staal                   12                    14                    26
## 20 Joe Thornt~                   8                    18                    26

                    # Making Graph

ggplot(data = d4a[1:5,],                          # Use the top five people from the d4a dataset
  aes(x = Name,                                  # X axis is for the names
  y = `Average Powerplay Points`,                  # Y axis is for average power play points
  fill = `Average Powerplay Goals`)) +              # Fill color with average power play goals
  geom_bar(stat = "identity", position= "dodge")  +
  labs(title = "Most Powerplay Goals, Assists and Points", # Set title and subtitle
  subtitle = "In each players average season",
  x = "Player",                              # Set x label and y label
  y = "Number of Points") +
  theme(axis.text.x = element_text(size = 8))          # Fix names spacing
```



#4b. What player has the most goals,assists and points in thier average Penatly Kill per season?

```
                    # Finding the Results

d4bID <- scoring_data %>%                         # Assign scoring data to new variable
 filter(lgID=="NHL")%>%                          # Filter data by players who are in the NHL
 group_by(playerID) %>%                           # Group the data by their player ID
 summarise(.groups = "drop",                       # Fix the ungrouping output error
 SHG = round(sum(SHG/n())),                         # Average goals will be the summation of their
goals divided by the # of seasons
 SHA = round(sum(SHA/n()))) %>%                          # Same as ^ but with assists instead
 mutate(SHP = SHA + SHG) %>%                          # New Column Called PPP which is PPA +
PPG
 arrange(desc(SHP)) %>%                           # Arrange by the most average points
 na.omit(d4bID)                           # Omit all Na's in dataset

                    # Displaying Names Instead of PlayerID
```

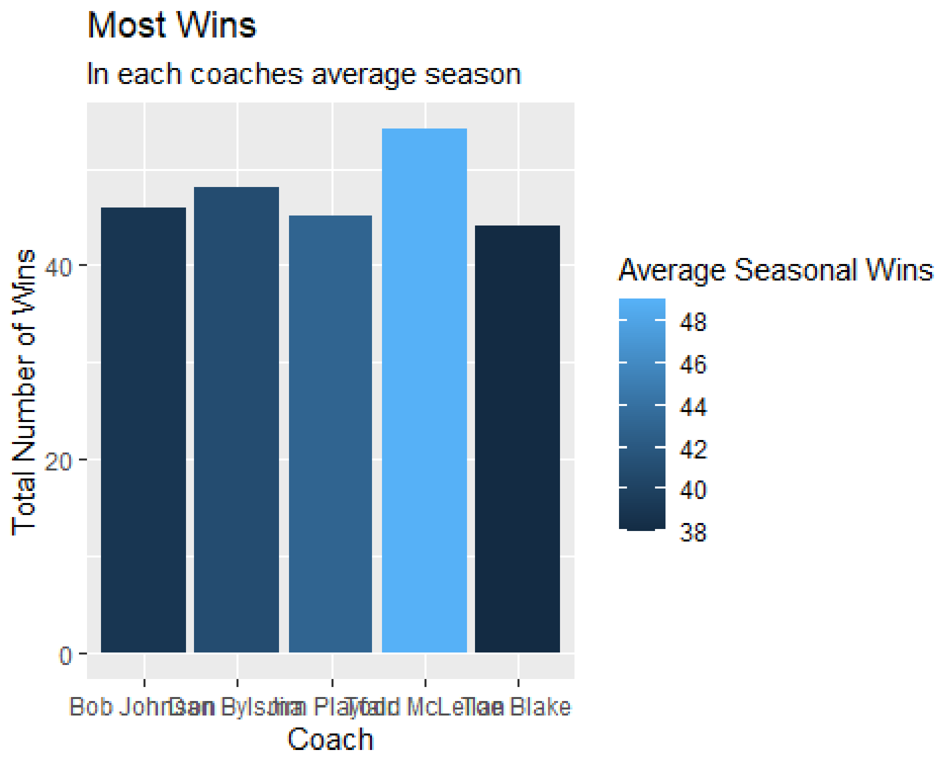
```
d4b <- left_join(d4bID,Players,"playerID") %>%          # Join together d4aID with Players
dataset to replace PlayerID with their names
  mutate(Name = paste(firstName,lastName)) %>%          # Join together the first and last
name in the Players
  select("Name" = Name,                               # Keep four variables
         "Average Shorthand Goals" = SHG,
         "Average Shorthand Assists" = SHA,
         "Average Shorthand Points" = SHP)

                    # Displaying Results

head(d4b,20)

## # A tibble: 20 x 4
##    Name         `Average Shorthand G~ `Average Shorthand A~ `Average Shorthand P~
##    <chr>                        <dbl>                 <dbl>                 <dbl>
##  1 Mike Richa~                      4                     1                     5
##  2 Pavel Bure                       3                     1                     4
##  3 Adam Henri~                      2                     2                     4
##  4 Michael Pe~                      2                     2                     4
##  5 Eric Perrin                      1                     3                     4
##  6 Jordan Sta~                      2                     2                     4
##  7 Daniel Alf~                      2                     1                     3
##  8 Jamie Benn                       2                     1                     3
##  9 Alexandre ~                      2                     1                     3
## 10 Andrew Cas~                      1                     2                     3
## 11 Erik Condra                      1                     2                     3
## 12 Sergei Fed~                      2                     1                     3
## 13 Theoren Fl~                      2                     1                     3
## 14 Marian Hos~                      2                     1                     3
## 15 Chris Kelly                      1                     2                     3
## 16 Anze Kopit~                      2                     1                     3
## 17 Ryan Malone                      2                     1                     3
## 18 Brad March~                      2                     1                     3
## 19 Rick Nash                        2                     1                     3
## 20 Ziggy Palf~                      2                     1                     3

                    # Making Graph

ggplot(data = d4b[1:5,],                              # Use the top five people from the d4b dataset
  aes(x = Name,                                       # X axis is for the names
  y = `Average Shorthand Points`,                     # Y asix is for average shorthand points
  fill = `Average Shorthand Goals`)) +                # Fill color with average shorthand goals
  geom_bar(stat = "identity",
  position= "dodge")  +
  labs(title = "Most Goals, Assists and Points",      # Set title and subtitle
  subtitle = "In each players average season",
  x = "Player",                              # Set x label and y label
  y = "Number of Points")
```



#5. What coaches has the most wins in their average season, post season and all time?

```
                    # Finding the Results

d5ID <- coaches_data %>%                        # Assign coaches data to new variable
 filter(lgID=="NHL")%>%                        # Filter data by players who are in the NHL
 group_by(coachID) %>%                         # Group the data by their coach ID
 summarise(.groups = "drop",                    # Fix the ungrouping output error
 W = round(sum(w/n())),                        # Average win will be the summation of their win
divided by the # of seasons
 postw = round(sum(postw/n()))) %>%                 # Same as ^ but with Post Season wins
instead
 mutate(ATW = W + postw) %>%                      # New Column Called ATW which is W +
PostW
 arrange(desc(ATW)) %>%                          # Arrange by the most average points
 na.omit(d5ID)                              # Omit all Na's in dataset

                    # Displaying Names Instead of coachID

d5 <- left_join(d5ID,Coaches,"coachID") %>%            # Join together d5ID with Coaches
dataset to replace CoachID with their names
```

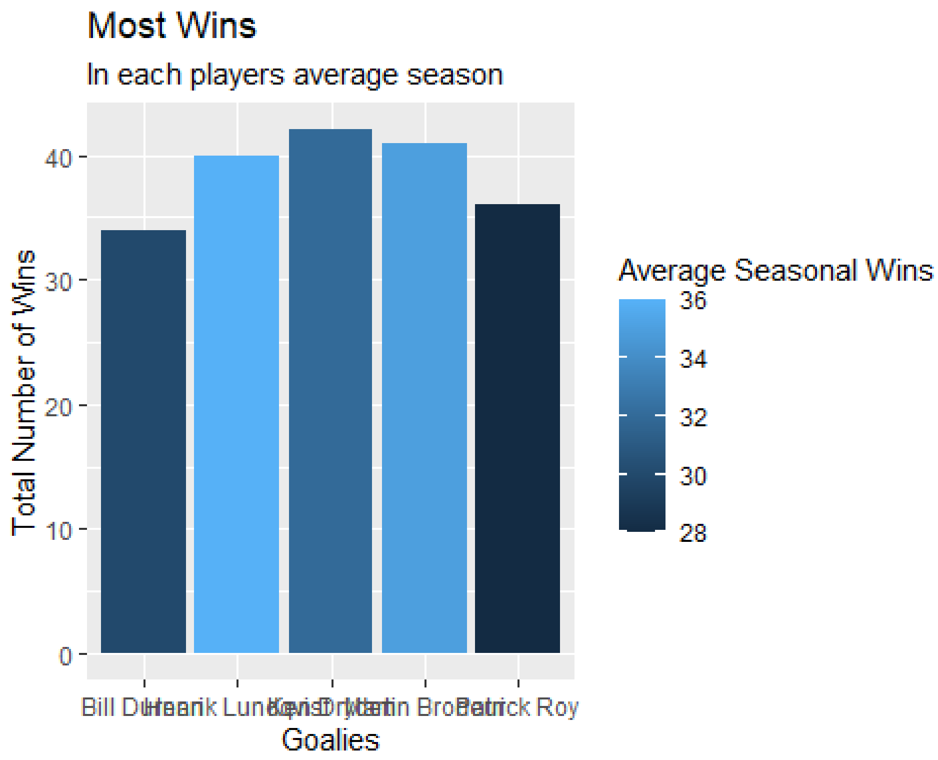
```
 mutate(Name = paste(firstName,lastName)) %>%          # Join together the first and last
name in the Players
 select("Name" = Name,                              # Keep four variables
       "Average Seasonal Wins" = W,
       "Average Post-Season Wins" = postw,
       "Average All Time Wins" = ATW)

                         # Displaying Results

head(d5,20)

## # A tibble: 20 x 4
##    Name        `Average Seasonal Wi~ `Average Post-Season ~ `Average All Time W~
##    <chr>                     <dbl>                 <dbl>                <dbl>
##  1 Todd McLel~                  49                     5                   54
##  2 Dan Bylsma                   41                     7                   48
##  3 Bob Johnson                  39                     7                   46
##  4 Jim Playfa~                  43                     2                   45
##  5 Toe Blake                    38                     6                   44
##  6 Paul MacLe~                  41                     3                   44
##  7 Terry O'Re~                  38                     6                   44
##  8 Kevin Dine~                  38                     3                   41
##  9 Bill Barber                  36                     2                   38
## 10 Mario Trem~                  36                     2                   38
## 11 Dale Hunter                  30                     7                   37
## 12 Cooney Wei~                  29                     5                   34
## 13 Kevin Lowe                   32                     1                   33
## 14 Billy Ingl~                  28                     1                   29
## 15 Keith Allen                  26                     2                   28
## 16 Dit Clapper                  26                     2                   28
## 17 Doug Harvey                  26                     2                   28
## 18 Frank Patr~                  24                     1                   25
## 19 Alex Curry                   24                     0                   24
## 20 Lou Lamori~                  17                     5                   22

                         # Making Graph

ggplot(data = d5[1:5,],                            # Use the top five people from the d5 dataset
 aes(x = Name,                                     # X axis is for the names
 y = `Average All Time Wins`,                      # Y asix is for average all time wins
 fill = `Average Seasonal Wins`)) +                # Fill color with average seasonal wins
 geom_bar(stat = "identity",
 position= "dodge")  +
 labs(title = "Most Wins",                         # Set title and subtitle
 subtitle = "In each coaches average season",
 x = "Coach",                                    # Set x label and y label
 y = "Total Number of Wins")
```



#6a. What goalie has the most wins in their average season, post season and all time?

```
                              # Finding the Results

d6aID <- goalies_data %>%                             # Assign goalies data to new variable
 filter(lgID=="NHL")%>%                              # Filter data by players who are in the NHL
 group_by(playerID) %>%                               # Group the data by their player ID
 summarise(.groups = "drop",                          # Fix the ungrouping output error
 W = round(sum(W/n())),                             # Average win will be the summation of their
win divided by the # of seasons
 PostW = round(sum(PostW/n()))) %>%                      # Same as ^ but with Post Season wins
instead
 mutate(ATW = W + PostW) %>%                         # New Column Called ATW which is W +
PostW
 arrange(desc(ATW))                             # Arrange by the most average points
                                                     # Omit all Na's in dataset

                              # Displaying Names Instead of playerID

d6a <- left_join(d6aID,Players,"playerID") %>%           # Join together d6aID with Players
dataset to replace PlayerID with their names
 mutate(Name = paste(firstName,lastName)) %>%          # Join together the first and last
```

```
name in the Players
  select("Name" = Name,                          # Keep four variables
         "Average Seasonal Wins" = W,
         "Average Post-Season Wins" = PostW,
         "Average All Time Wins" = ATW)

                         # Displaying Results

head(d6a,20)

## # A tibble: 20 x 4
##    Name        `Average Seasonal Wi~ `Average Post-Season~ `Average All Time W~
##    <chr>                      <dbl>                 <dbl>                <dbl>
##  1 Ken Dryden                    32                    10                   42
##  2 Martin Brod~                  35                     6                   41
##  3 Henrik Lund~                  36                     4                   40
##  4 Patrick Roy                   28                     8                   36
##  5 Bill Durnan                   30                     4                   34
##  6 Marc-Andre ~                  28                     5                   33
##  7 Evgeni Nabo~                  28                     4                   32
##  8 Cam Ward                      29                     3                   32
##  9 Roberto Luo~                  28                     3                   31
## 10 Ryan Miller                   28                     3                   31
## 11 Ed Belfour                    25                     5                   30
## 12 Roman Cechm~                  28                     2                   30
## 13 Miikka Kipr~                  28                     2                   30
## 14 Antti Niemi                   24                     6                   30
## 15 Jonathan Qu~                  26                     4                   30
## 16 Tony Esposi~                  26                     3                   29
## 17 Frank Brims~                  25                     3                   28
## 18 Dominik Has~                  24                     4                   28
## 19 Tim Thomas                    24                     4                   28
## 20 Niklas Back~                  27                     0                   27

                         # Making Graph

ggplot(data = d6a[1:5,],                          # Select top 5 people from d6a dataset
  aes(x = Name,                                   # x axis is for the names
  y = `Average All Time Wins`,                    # Y axis is for average all time wins
  fill = `Average Seasonal Wins`)) +              # Fill color with average seasonal wins
  geom_bar(stat = "identity",
  position= "dodge")  +
  labs(title = "Most Wins",          # Set title and subtitle
  subtitle = "In each players average season",
  x = "Goalies",                                  # Set x label and y label
  y = "Total Number of Wins")
```



#6b. What experienced goalies have the best Save percent of all time?

```
                    # Finding the Results

d6bID <- goalies_data%>%                              # Assign goalies data to new variable
 drop_na(SA) %>%                                 # Drop all the Na's
 filter(lgID=="NHL") %>%                             # Filtered only NHL players
 group_by(playerID) %>%                             # Grouped by playerID
 summarise(.groups = "drop",                          # Fix the ungrouping output error
 GP=sum(GP),                                # Sum up the games played
 GA=sum(GA),                                # Sum up goals against
 SA=sum(SA),                                # Sum up shots against
 SV = round((1-GA/SA)*100,2)) %>%                     # Create the percent saved
 filter(GP>500)%>%                               # Filtered by games played to get the most
expeirenced goalies
 arrange(desc(SV))                               # Arrange in descending value

                    # Displaying Names Instead of playerID

d6b <- left_join(d6bID,Players,"playerID") %>%          # Join together d6bID with Players
dataset to replace PlayerID with their names
 mutate(Name = paste(firstName,lastName)) %>%          # Join together the first and last
name in the Players
 select("Name" = Name,                            # Keep three variables
     "Games Played" = GP,
    "Save Percent (%)" = SV)

                    # Displaying Results

head(d6b,20)

## # A tibble: 20 x 3
##    Name                 `Games Played` `Save Percent (%)`
##    <chr>                         <dbl>              <dbl>
##  1 Dominik Hasek                   735               92.2
##  2 Roberto Luongo                  727               91.9
##  3 Tomas Vokoun                    680               91.7
##  4 Miikka Kiprusoff                599               91.4
##  5 Martin Brodeur                 1191               91.3
##  6 Jean-Sebastien Giguere          557               91.3
##  7 Evgeni Nabokov                  605               91.2
##  8 Patrick Roy                    1029               91.0
##  9 Marty Turco                     543               91.0
## 10 Jose Theodore                   633               90.9
## 11 Dwayne Roloson                  606               90.8
## 12 Nikolai Khabibulin              783               90.7
## 13 Ed Belfour                      963               90.6
## 14 Olaf Kolzig                     719               90.6
## 15 Curtis Joseph                   943               90.6
## 16 Felix Potvin                    635               90.5
## 17 Chris Osgood                    744               90.5
## 18 Tommy Salo                      526               90.5
## 19 Mike Richter                    666               90.4
## 20 Jocelyn Thibault                586               90.4
```

#7. What experienced player took the least amount of penalites?

```
                    # Finding the Results

d7ID <- scoring_data %>%                        # Use scoring data
  group_by(playerID) %>%                        # Group by playerID
  filter(lgID=="NHL",pos != "G") %>%                # Show only NHL players and non goalies
  summarise(.groups = "drop",                   # Fix the ungrouping output error
  PIM=sum(PIM),                                # Penalty minutes
  GP=sum(GP))%>%                                # Games played
  filter(GP>750) %>%                            # Players have to play at least 750 games
  arrange(PIM)                               #arrange by PIM

                    # Displaying Names Instead of playerID

d7 <- left_join(d7ID,Players,"playerID") %>%          # Join together d7ID with Players
dataset to replace PlayerID with their names
```

```
 mutate(Name = paste(firstName,lastName)) %>%          # Join together the first and last
name in the Players
 select("Name" = Name,                                # Keep three variables
       "Games Played" = GP,
       "Penalty Minutes" = PIM)

                    # Displaying Results

head(d7,20)

## # A tibble: 20 x 3
##    Name              `Games Played` `Penalty Minutes`
##    <chr>                      <int>             <int>
##  1 Val Fonteyne                 820                26
##  2 Bill Quackenbush             774                95
##  3 Woody Dumart                 772                99
##  4 Butch Goring                1107               102
##  5 Dave Keon                   1296               117
##  6 Robert Kron                  771               119
##  7 Rick Kehoe                   906               120
##  8 Don Marshall                1176               127
##  9 Phil Goyette                 941               131
## 10 Mikael Andersson             761               134
## 11 Fred Stanfield               914               134
## 12 Harry Watson                 809               150
## 13 Jody Hull                    831               156
## 14 Rick Middleton              1005               157
## 15 Mark Napier                  767               157
## 16 Jay Pandolfo                 881               162
## 17 Craig Janney                 760               170
## 18 Sami Kapanen                 831               175
## 19 Peter McNab                  954               179
## 20 Brad Richards                854               199
```

#8. Who are the greatest players of all time based off of Awards they recieved? (Noah)

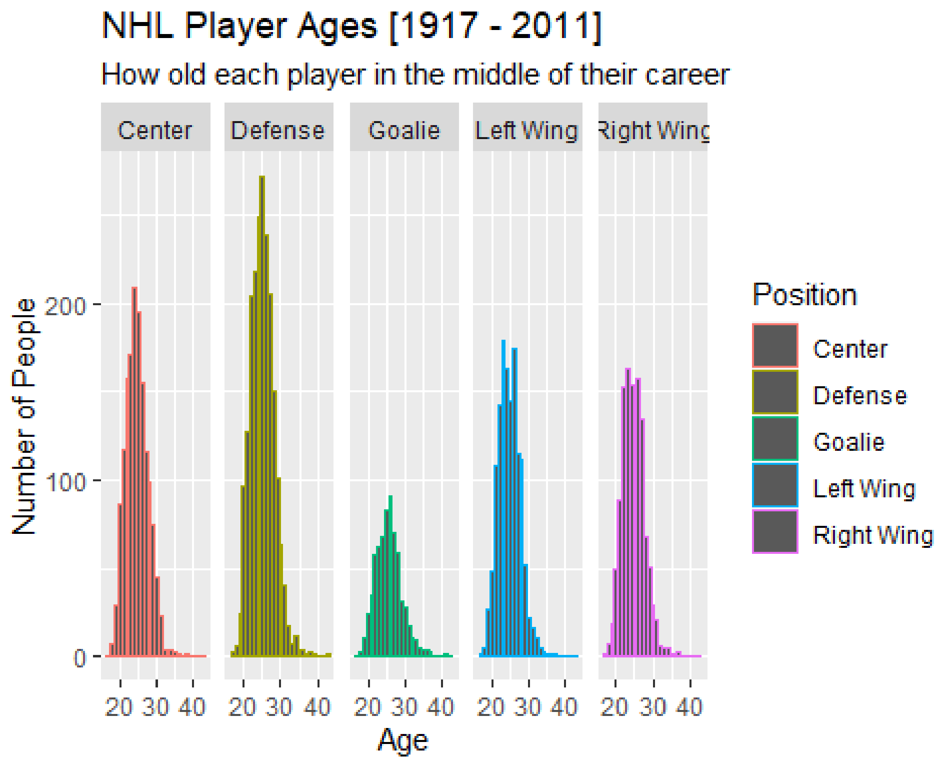
```
                    # Finding the Results

d8ID <- awards_players_data %>%                    # Assign awards_player_data to new
variable
 group_by(playerID) %>%                           # Group by their player ID
 filter(lgID == "NHL") %>%                        # Filter for those who are in the NHL
 summarise(.groups = 'drop',                      # Fix the ungrouping output error
 Number_of_Awards = sum(n())) %>%                      # "#" of awards is the sum of awards
given to a player
 arrange(desc(Number_of_Awards))                  # Arrange by highest amount of awards
```

```
                        # Displaying Names Instead of playerID

d8 <- left_join(d8ID,Players,"playerID") %>%             # Join together d8ID with Players
dataset to replace the playerIDs with their names
  mutate(Name = paste(firstName,lastName)) %>%            # Join together the first and last
name in the Players
  select("Name" = Name,                              # Keep two variables
       "Number of Awards"=Number_of_Awards)

                        # Display Results

head(d8,50)
```

```
## # A tibble: 50 x 2
##    Name           `Number of Awards`
##    <chr>                       <int>
##  1 Wayne Gretzky                  49
##  2 Gordie Howe                    33
##  3 Mario Lemieux                  28
##  4 Raymond Bourque                26
##  5 Bobby Orr                      26
##  6 Nicklas Lidstrom               21
##  7 Dominik Hasek                  20
##  8 Martin Brodeur                 18
##  9 Doug Harvey                    18
## 10 Bobby Hull                     18
## # ... with 40 more rows
```

#9. What are the ages of NHL players in the middle of their career? Find the standard deviation or third quartile for each position.

```
                        # Finding the Results

d9ID <- master_data %>%                              # Assign Master Data to new variable
  select(playerID,                                  # Only keep PlayerID, firstNHL, LastNHL
  firstNHL,                                        # Birthyear and position
  lastNHL,
  birthYear,
  pos,) %>%
  na.omit(firstNHL,lastNHL) %>%                         # Omit all Na values
  mutate(
  Age = round((firstNHL + lastNHL)*.5) - birthYear) %>%     # Find age by taking their
average NHL career and subtract by their birthyear
  select(playerID,                                  # Only keep PlayerID, Age, and Position
  Age,
  pos) %>%
  filter(pos != "D/L",                               # Get rid of outliers
 pos != "F",
 pos != "L/D",
 pos != "L/C" )

                    # Find the 3rd Quartile (The average 3rd quartile is 27)

AgeID <- d9ID %>%                               # Assign d9ID to new variable
 group_by(pos) %>%                               # Group by their position
 summarise(.groups = "drop",                         # Fix the ungrouping output error
 "3rd Quartile" = quantile(Age))                     # Find Quartile stats
AgeID <- AgeID[c(4,9,14,19,24),]                      # Only Keep the third Quartile



d9ID$pos[d9ID$pos == "C"] <- "Center"                    # Replace C with Center
d9ID$pos[d9ID$pos == "D"] <- "Defense"                   # Replace D with Defense
d9ID$pos[d9ID$pos == "L"] <- "Left Wing"                 # Replace L with Left Wing
d9ID$pos[d9ID$pos == "R"] <- "Right Wing"                 # Replace R with Right Wing
d9ID$pos[d9ID$pos == "G"] <- "Goalie"                    # Replace G with Goalie


ggplot(d9ID,aes(Age)) +                               # Create ggplot from d9ID with AES age
geom_histogram(binwidth = 1,aes(color = pos)) +           # Make it a histogram with
bandwidth 1 and color based off position
facet_wrap(~pos)+                                  # Make separate graphs for each position
facet_grid(~pos) +                                 # Make the graphs side by side
labs(title = "NHL Player Ages [1917 - 2011]",             # Create title and subtitle
subtitle = "How old each player in the middle of their career",
y = "Number of People",                            # Set x,y title
x = "Age", color = "Position") +                       # set legend to be position
scale_x_continuous(breaks = seq(0, 60, by = 10))          # Set frequency of ticks
```



#10 Who would we want on our team

```
                    # Finding Final Results

d10PID <- full_join(d1ID[,1],d2aID[,1],"playerID")        # d10PID (The P stands for players)
d10PID <- full_join(d10PID,d2bID[,1],"playerID")         # Get each playerID that was top 50
for any category
d10PID <- full_join(d10PID,d3ID[,1],"playerID")          # Only keep the playerID
d10PID <- full_join(d10PID,d4aID[,1],"playerID")
d10PID <- full_join(d10PID,d4bID[,1],"playerID")
d10PID <- full_join(d10PID,d7ID[,1],"playerID")
d10PID <- full_join(d10PID,d8ID[,1],"playerID")

d10PID <- full_join(d10PID,d1ID[1:50,1:2],"playerID")       # Find the stats of each player in
the list made before^
d10PID <- full_join(d10PID,d2aID[1:50,1:2],"playerID")      # If the player did now make top
50 for a category we will be making the na
d10PID <- full_join(d10PID,d2bID[1:50,1:2],"playerID")      # into a 0. If they did we will be
making it into a 1.
d10PID <- full_join(d10PID,d3ID[1:50,1:2],"playerID")       # We will tally up the stats for each
player to see
d10PID <- full_join(d10PID,d4aID[1:50,1:2],"playerID")      # Which good players were the
most balanced.
d10PID <- full_join(d10PID,d4bID[1:50,1:2],"playerID")      # We favor those who are top 50
in multiple categories rather then
```

```
d10PID <- full_join(d10PID,d7ID[1:50,1:2],"playerID")      # Those who are only number one
in a category
d10PID <- full_join(d10PID,d8ID[1:50,1:2],"playerID")
d10PID[,2:9][!is.na(d10PID[,2:9])] <- 1                 # Make one if they are in top 50 for each
category
d10PID[,2:9][is.na(d10PID[,2:9])] <- 0                  # Make zero if they are not in top 50

d10PID <- cbind(d10PID, "Top" = rowSums(d10PID[,2:9])) %>%  # Use cbind to sum up the
rows of ones for each player
select(playerID,"Top")                              # Only keep the player ID and the summation of
the Top 50s

d10PID <- left_join(d10PID,d9ID,"playerID")             # Combine the dataset with the file that
has their ages and position

                         # Best/Balanced Players
# Centers
d10CID <- d10PID %>%                                # Assign d10pID to d10CID (C stands for center)
  filter(pos == "Center", Age <= 27) %>%                # Only use centers and those ages of 27
and lower (Found in part 9)
  arrange(desc(Top)) %>%                              # Arrange by the most top 50
  head(3) %>%                                     # Only keep Top 3
  select(playerID,pos)                              # Only keep the variables playerID and Position

# Left Wings
d10LWID <- d10PID %>%                               # Assign d10pID to d10LWID (LW stands for
Left Wing)
  filter(pos == "Left Wing", Age <= 27) %>%             # Only use centers and those ages of 27
and lower (Found in part 9)
  arrange(desc(Top)) %>%                              # Arrange by the most top 50
  head(3) %>%                                     # Only keep Top 3
  select(playerID,pos)                              # Only keep the variables playerID and Position

# Right Wings
d10RWID <- d10PID %>%                               # Assign d10pID to d10RWID (RW stands for
Right Wing)
  filter(pos == "Right Wing", Age <= 27) %>%            # Only use centers and those ages of 27
and lower (Found in part 9)
  arrange(desc(Top)) %>%                              # Arrange by the most top 50
  head(3) %>%                                     # Only keep Top 3
  select(playerID,pos)                              # Only keep the variables playerID and Position

# Defense
d10DID <- d10PID %>%                                # Assign d10pID to d10DID (D stands for
Defense)
  filter(pos == "Defense", Age <= 27) %>%               # Only use centers and those ages of 27
```

```
and lower (Found in part 9)
  arrange(desc(Top)) %>%                    # Arrange by the most top 50
  head(6) %>%                           # Only keep Top 6
  select(playerID,pos)                       # Only keep the variables playerID and Position

# First Goalie
d10G1ID <- d6aID[1,1]                        # Find the goalie with the most wins
d10G1ID[1,2]<- "Goalie"                      # Assign Position to Goalie

# Second Goalie
d10G2ID <- d6bID[1,1]                        # Find the goalie with the highest save Percent
d10G2ID[1,2]<- "Goalie"                      # Assign Position to Goalie

# Coach
d10ID <- d5[1,1]                          # Find the best coach


# Team Roster
d10 <- d10CID[1:3,1:2]                       # Combine the data into one team Roster
d10[4:6,1:2] <- d10LWID[1:3,1:2]
d10[7:9,1:2] <- d10RWID[1:3,1:2]
d10[10:15,1:2] <- d10DID[1:6,1:2]
d10[16,1:2] <- d10G1ID
d10[17,1:2] <- d10G2ID

# Replace all player IDs with their actual names
d10 <- left_join(d10,Players,"playerID") %>%          # Combine players with d10
  mutate(Name = paste(firstName,lastName)) %>%          # Make a name column
  select("Name" = Name, "Position" = pos)           # Only keep their name and Position

# Coach Roster
d10[18,1] <- d10ID                         # Add coach to roster
d10[18,2] <- "Coach"

# Display results
head(d10,18)

##                 Name   Position
## 1       Evgeni Malkin     Center
## 2   Nicklas Backstrom     Center
## 3        Anze Kopitar     Center
## 4       Alex Ovechkin  Left Wing
## 5   Henrik Zetterberg  Left Wing
## 6      Ilya Kovalchuk  Left Wing
## 7          Mike Bossy Right Wing
## 8      Theoren Fleury Right Wing
## 9        Dany Heatley Right Wing
## 10       Denis Potvin    Defense
## 11          Bobby Orr    Defense
## 12  Scott Niedermayer    Defense
## 13        Behn Wilson    Defense
## 14       Dion Phaneuf    Defense
## 15 John-Michael Liles    Defense
## 16         Ken Dryden     Goalie
## 17      Dominik Hasek     Goalie
## 18      Todd McLellan      Coach
```