ggplot2 Assignment

Adam Kiehl

2022-09-14

Data

```
# read in the data from a .csv file
data <- read.csv('players.csv') %>%
  # calculate average time on ice
  mutate(ATOI = round(TOI / GP, 3)) %>%
  # convert position to a factor
  mutate(Pos = as.factor(Pos)) %>%
  # remove empty rows
  filter(Player != '') %>%
  # remove non-standard positions
  filter(!(Pos %in% c('F', 'W', 'G')))

head(data)
```

```
##               Player Age Team Pos GP G  A PTS PlusMin  PS  S TOI BLK HIT   ATOI
## 1 Nicholas Abruzzese  22  TOR   C  9 1  0   1      -1 0.0  8  92   3   7 10.222
## 2       Noel Acciari  30  FLA   C 20 3  5   8       2 0.6 32 240  16  48 12.000
## 3      Calen Addison  21  MIN   D 15 2  2   4      -4 0.3 17 207   6  12 13.800
## 4    Andrew Agozzino  31  OTT  LW  1 0  0   0       0 0.0  1   7   0   4  7.000
## 5         Jack Ahcan  24  BOS   D  6 1  0   1      -3 0.1  5  96   5   8 16.000
## 6      Sebastian Aho  25  NYI   D 36 2 10  12      -6 1.7 34 592  42  32 16.444
```

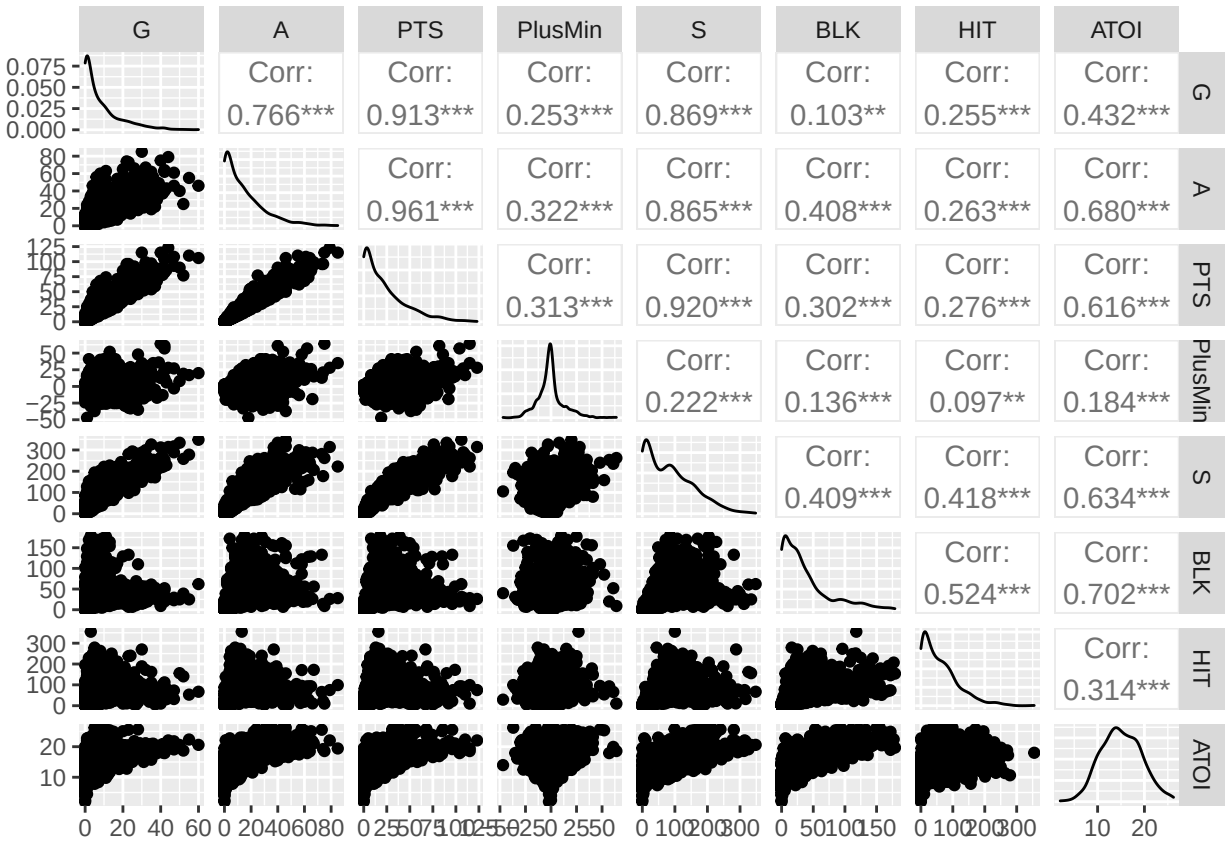
Exploratory Data Analysis

```
# examine characteristics of data
ExpData(data)
```

```
##                                       Descriptions     Value
## 1                                Sample size (nrow)       959
## 2                           No. of variables (ncol)        15
## 3                   No. of numeric/interger variables        12
## 4                             No. of factor variables         1
## 5                               No. of text variables         2
## 6                            No. of logical variables         0
## 7                         No. of identifier variables         1
## 8                               No. of date variables         0
## 9            No. of zero variance variables (uniform)         0
```

```
## 10                %. of variables having complete cases 100% (15)
## 11   %. of variables having >0% and <50% missing cases    0% (0)
## 12 %. of variables having >=50% and <90% missing cases    0% (0)
## 13          %. of variables having >=90% missing cases    0% (0)
```

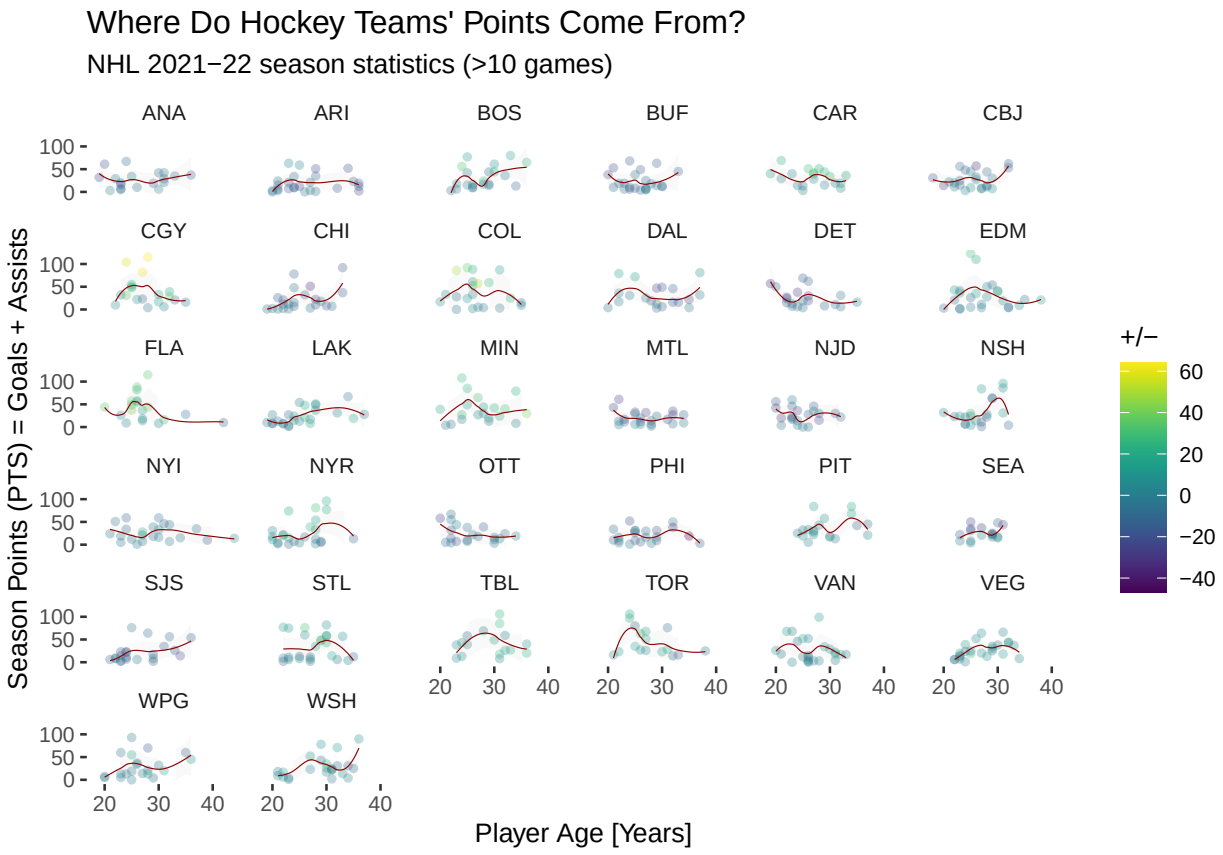
```
# explore structure of data
data %>%
  select(-c(Player, Age, Team, Pos, GP, TOI, PS)) %>%
  ggpairs(progress = FALSE)
```



Player Ages Between Teams

```
plt1 <- data %>%
  # remove players without an identified team
  filter(Team != 'TOT') %>%
  # remove players with <10 games
  filter(GP > 10) %>%
  # create aesthetic mapping
  ggplot(mapping = aes(x = Age,
                       y = PTS,
                       color = PlusMin)) +
  # add scatter plot
  geom_point(alpha = .3, size = 1) +
  # add trend lines
```

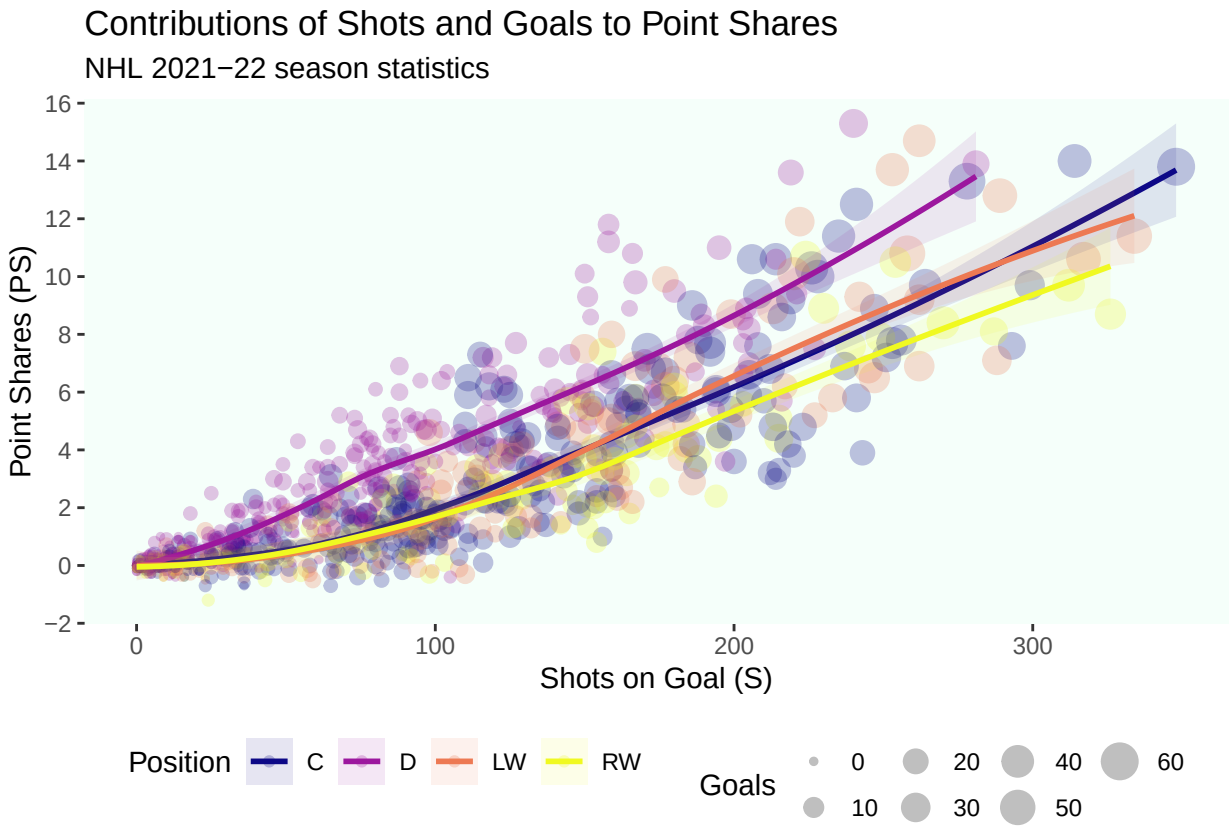
```
  geom_smooth(method = 'loess', alpha = .05, size = .2, col = 'red4') +
  # plot separately by team
  facet_wrap(. ~ Team, shrink = FALSE) +
  # format axes
  scale_x_continuous(breaks = seq(20, 40, 10)) +
  scale_y_continuous(breaks = seq(0, 100, 50),
                     limits = c(-3, max(data$PTS))) +
  # change legend format and title
  scale_color_continuous(breaks = seq(-50, 50, 25)) +
  guides(color = guide_colorbar(title = '+/-')) +
  # change plot theme
  theme_tufte(base_size = 10, base_family = 'sans') +
  # change color palette
  scale_color_viridis() +
  # add plot and axis titles
  labs(title = 'Where Do Hockey Teams\' Points Come From?',
       subtitle = 'NHL 2021-22 season statistics (>10 games)',
       x = 'Player Age [Years]',
       y = 'Season Points (PTS) = Goals + Assists'); plt1
```




```
pdf('plot1.pdf', height = 4, width = 8)
print(plt1)
invisible(dev.off())
```

Point Shares

```
plt2 <- ggplot(data = data,
               # create aesthetic mapping
               mapping = aes(x = S,
                             y = PS,
                             color = Pos,
                             fill = Pos,
                             group = Pos)) +
  # add scatter plot
  geom_point(aes(size = G),
             alpha = .25) +
  # add trend lines
  geom_smooth(method = 'loess', alpha = .1) +
  # format y axis
  scale_y_continuous(breaks = seq(-2, 16, 2)) +
  # change plot theme
  theme_tufte(base_size = 11, base_family = 'sans') +
  # change legend titles
  guides(size = guide_legend(title = 'Goals'),
         fill = guide_legend(title = 'Position'),
         color = guide_legend(title = 'Position')) +
  # move legend to bottom of plot
  theme(legend.position = 'bottom',
        # change color of plot background
        panel.background = element_rect(fill = 'mintcream',
                                        color = 'mintcream')) +
  # change color palette for scatter plot and trend lines
  scale_color_viridis(discrete = TRUE, option = 'C') +
  scale_fill_viridis(discrete = TRUE, option = 'C') +
  # add plot and axis titles
  labs(title = 'Contributions of Shots and Goals to Point Shares',
       subtitle = 'NHL 2021-22 season statistics',
       x = 'Shots on Goal (S)',
       y = 'Point Shares (PS)'); plt2
```

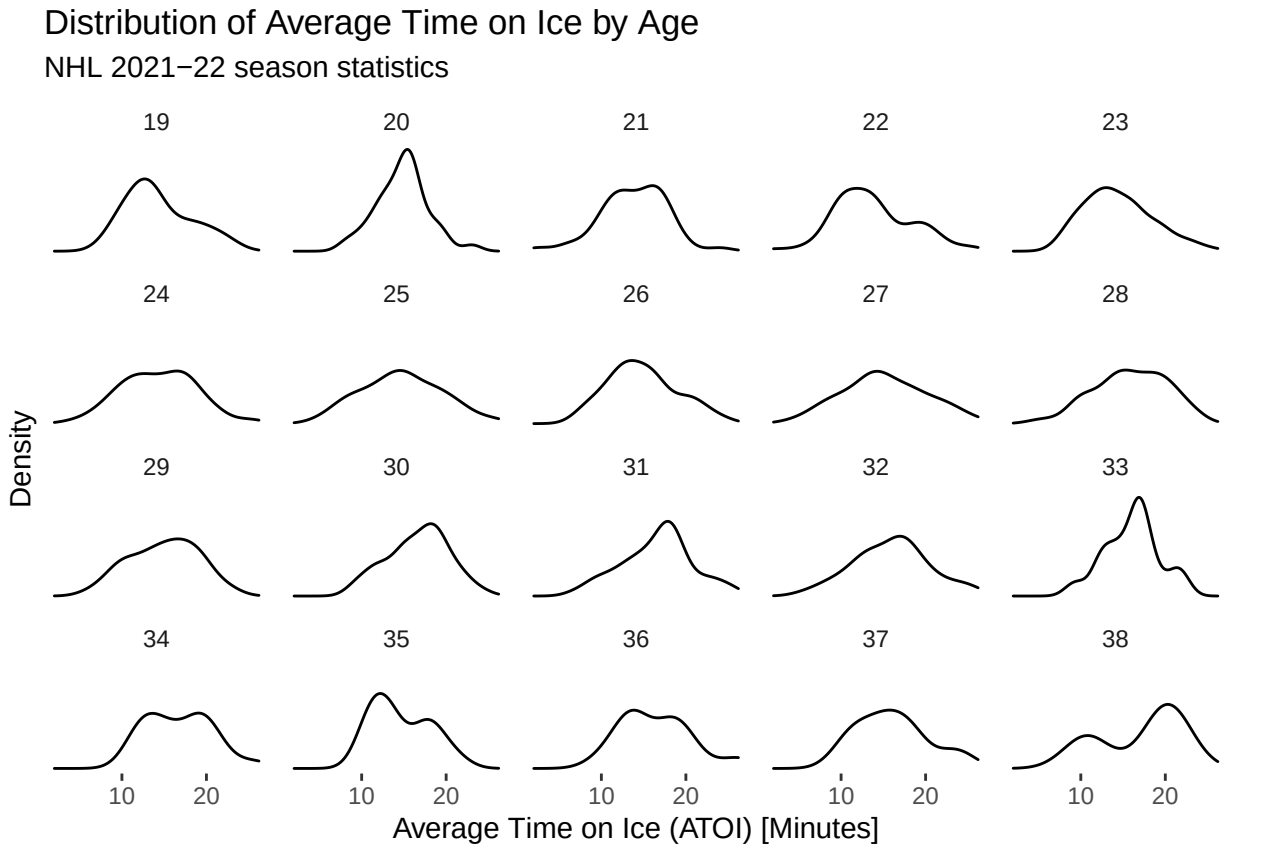



```
pdf('plot2.pdf', height = 4, width = 8)
print(plt2)
invisible(dev.off())
```

Average Time on Ice

```
plt3 <- data %>%
  merge((data %>%
           group_by(Age) %>%
           summarize(n = n())),
        by.x = 'Age',
        by.y = 'Age',
        all.x = TRUE) %>%
  # filter for age groups with >1 player
  filter(n > 1) %>%
  select(-n) %>%
  group_by(Age, Pos) %>%
  # create aesthetic mapping
  ggplot(mapping = aes(x = ATOI)) +
  # add density plots
  geom_density() +
  # plot separately by age
  facet_wrap(. ~ Age) +
  # remove breaks on y axis
```

```
  scale_y_continuous(breaks = c()) +
  # change plot theme
  theme_tufte(base_size = 11, base_family = 'sans') +
  # add plot and axis titles
  labs(title = 'Distribution of Average Time on Ice by Age',
       subtitle = 'NHL 2021-22 season statistics',
       x = 'Average Time on Ice (ATOI) [Minutes]',
       y = 'Density'); plt3
```




```
pdf('plot3.pdf', height = 4, width = 8)
print(plt3)
invisible(dev.off())
```