**Visualizing Professional Woman’s Hockey Data**

The Professional Woman’s Hockey League (PWHL) began its inaugural season in 2023-24. The league has players from 11 different countries. The league looks to expand its exposure and gain new fans for the future of the sport.

We will be investigating the player statistics from the league’s inaugural season. Our focus will be on all players including goalies which includes 147 athletes. Also, to discover which age groups and positions have had the most impact on the number of goals scored per game by each player.

In general, the main goal of the forwards is to stay in three different lanes across the ice, moving the puck between them to make the goalie move and open scoring opportunities. The defense players compliment the forwards by positioning themselves along the boundary of the offensive zone to prevent the opposing team from moving the puck away from the zone and provide more opportunities for the forwards to score. The goalie’s focus is to guard their team’s goal by positioning themselves in front of it to prevent the opposing team from scoring.

The different age groups represent a blend of experience and athleticism at a point in the player’s career. More experience should help the player score more goals because they would have more knowledge of the game and ideas on how to score. However, more experience comes with more aging and players with more experience may be past their years of peak athleticism. That is why our goal is to find if there is a perfect blend between the two (i.e. an ideal age and position group).

**Data Description**

**PWHL\_Final** - Name of the data set

**G\_Per\_GP** -Number of goals scored by the player per game played

**P** - Number of points scored

**Pos** – Position of the player (either Defense (D), Forward (F), or Goalie (G))

**Age** – Age of the player in years

The density plot below displays the distribution of goals per game played for each position. Use it to answer the following two questions.

A graph with different colored lines

Description automatically generated

1. What would you need to add to the R code below to add a title and change the x-axis label to "Goals Per Game Played"?

```{r}

ggplot(data = PWHL\_Final) +

geom\_density(aes(G\_Per\_GP, color = Pos, fill = Pos), alpha = 0.25) +

theme\_minimal()

```

+ labs(x = "Goals Per Game Played", title = "Goals between different Age Groups and Positions")

1. Describe the distribution of the Forward and Defense positions in the density plot above. Make sure to mention shape and skew. Give a possible reason why the Goalie’s curve is flat in this visual.

Key Points: Mention bell-shape and right skew for the forwards and defense, note the differences between the two. Understand why a Goalie’s curve would be flat.

Sample Answer:

The distribution of the goals per game played for Goalies is flat because they don’t score goals so there isn’t variation between goalies which creates that flat curve. The distribution of the goals per game played for the defense position has a steep, bell-shaped curve and is skewed right, with a center closer to 0 compared to the distribution for forwards. The forward’s distribution is bell-shaped and less steep than the ones for defense with its center to the right of the defense distribution. It is also skewed right.

1. Brainstorm some ideas on which combinations of age and positions would be ideal for a player to be to maximize their number of goals per game played. Do you see any significant differences among the different age groups? Explain.

A combination of the Forward position and any age group would suffice because there seems to be no significant differences between the three.

1. Let’s put this all together on one graph to see if it’s any easier to see the trends between the age groups and positions. First, create three different age groups “youngest” (22-25), “middle” (26-30), and “oldest” (31-36). Then, create side-by-side boxplots that display the range of goals per game for each group. Only include Forward and Defense positions. Make sure to add a theme, flip the axis, and change the x and y axis labels. Refer to the data description on the first page.

```{r}

PWHL\_Graph <- PWHL\_Final %>%

mutate(Age\_Group = case\_when(

Age >= 22 & Age <= 26 ~ "youngest",

Age >= 27 & Age <= 30 ~ "middle",

Age >= 31 & Age <= 36 ~ "oldest"

)) %>%

filter(Pos != "G")

PWHL\_Graph <- PWHL\_Graph %>%

mutate(Age\_Group = factor(Age\_Group, levels = c("youngest", "middle", "oldest")))

```

```{r}

ggplot(data = PWHL\_Graph, aes(x = Age\_Group, y = G\_Per\_GP, color = Pos)) +

geom\_boxplot() +

coord\_flip() +

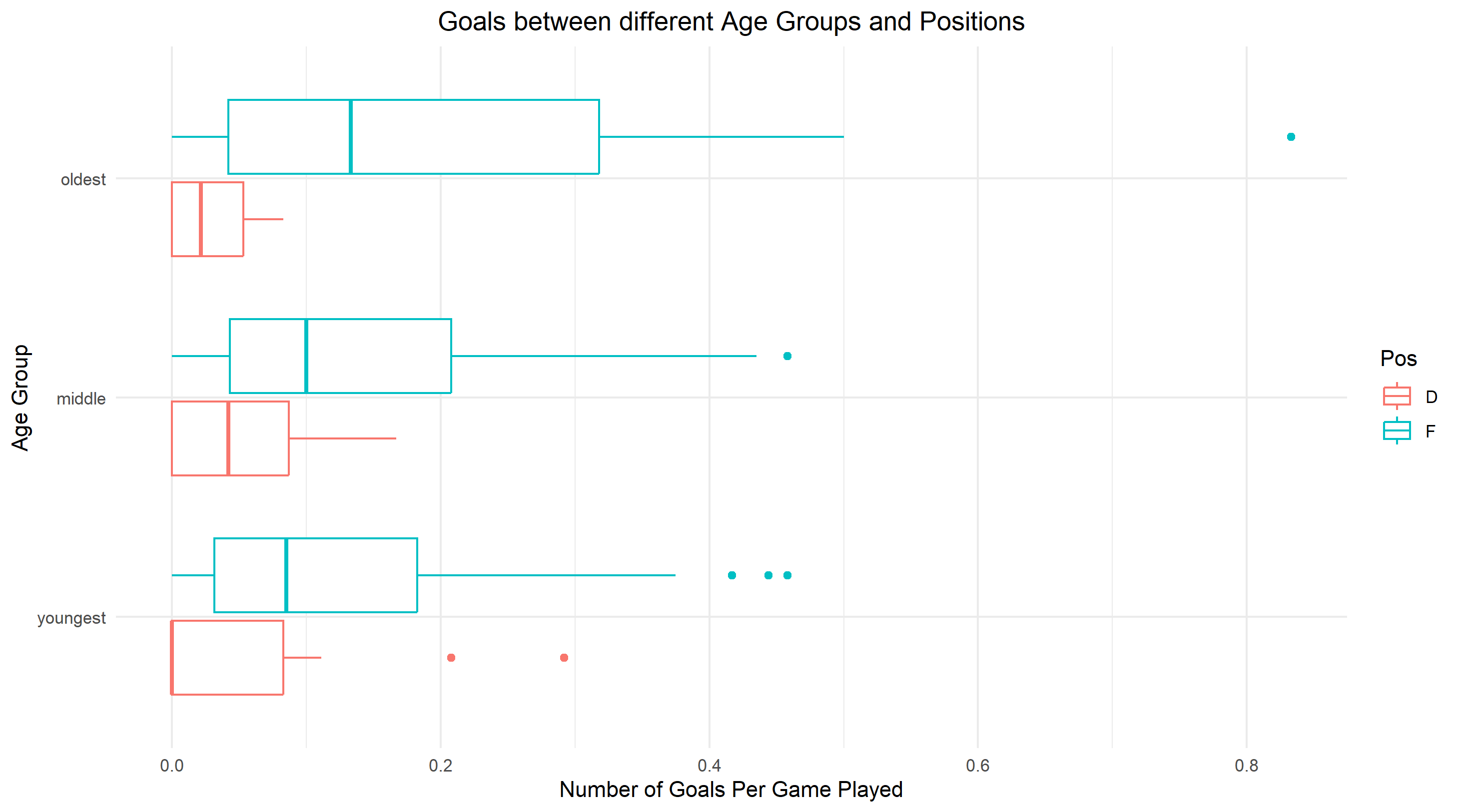
labs(x = "Age Group", y = "Number of Goals per Game Played", title = "Goals Per Game Played between different Age Groups and Positions") +

theme\_minimal() +

theme(plot.title = element\_text(hjust = 0.5))

```

Sample Graph:



1. What trends, if any, do you see in your graph? Does it confirm your original thoughts in question 3? What seems to be the ideal position and age group to maximize goals per game played?

For forwards, as age increases, so does the number of goals per game played. For defense, the middle age has the greatest number of goals per game played with oldest behind them and youngest with the least. The ideal position and age group is Forward and oldest.

1. Reread the summary on the first page and find a limitation to this dataset.

There has only been one season, so the data is limited to a small sample size. When the league has a few more seasons the data will likely be clearer and reflect what is going on better.