**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.

**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 = "Points 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. 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 points 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 = P, color = Pos)) +

geom\_boxplot() +

coord\_flip() +

labs(x = "Age Group", y = "Number of Points", title = "Points between different Age Groups and Positions") +

theme\_minimal()

```