

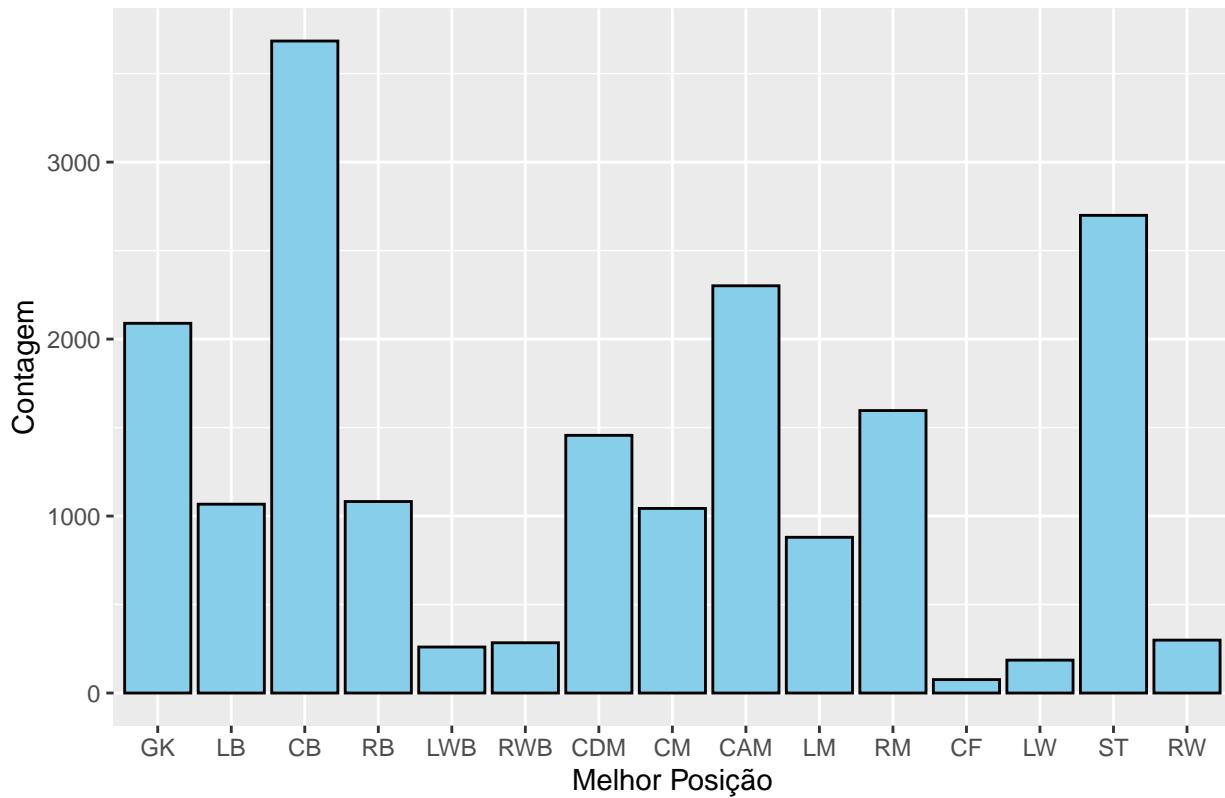
Luiz Henrique Barretta Francisco - GRR20213026

Análise de Dados FIFA 21

```
df <- read_csv("players_dataset.csv")
ordem_posicoes <- c("GK", "LB", "CB", "RB", "LWB", "RWB", "CDM", "CM", "CAM", "LM",
                     "RM", "CF", "LW", "ST", "RW")
df$Best_position <- factor(df$Best_position, levels = ordem_posicoes)

hist_best_position <- ggplot(df, aes(x = Best_position)) +
  geom_bar(fill = "skyblue", color = "black") +
  labs(title = "Distribuição das Melhores Posições",
       x = "Melhor Posição", y = "Contagem")
print(hist_best_position)
```

Distribuição das Melhores Posições



Acima, temos a distribuição das melhores posições dos jogadores da base,

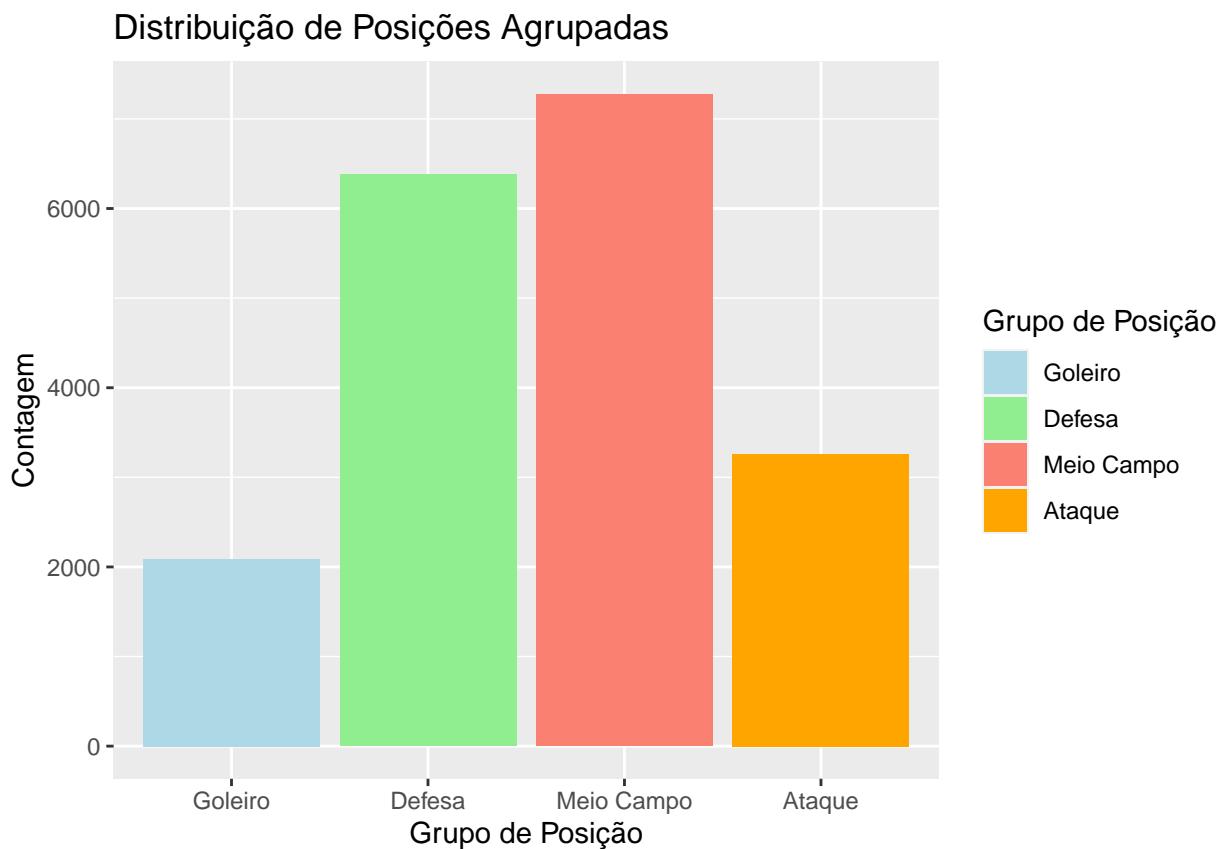
```
df <- df %>%
  mutate(position_group = case_when(
    Best_position %in% c("GK") ~ "Goleiro",
    Best_position %in% c("LB", "CB", "RB", "LWB", "RWB") ~ "Defesa",
    Best_position %in% c("CDM", "CM", "CAM", "LM", "RM") ~ "Meio Campo",
```

```

Best_position %in% c("CF", "LW", "ST", "RW") ~ "Ataque"))

df$position_group <- factor(df$position_group, levels =
                           c("Goleiro", "Defesa", "Meio Campo", "Ataque"))
hist_best_position2 <- ggplot(df, aes(x = position_group, fill = position_group)) +
  geom_bar() +
  scale_fill_manual(values = c("Goleiro" = "lightblue", "Defesa" = "lightgreen",
                               "Meio Campo" = "salmon", "Ataque" = "orange")) +
  labs(title = "Distribuição de Posições Agrupadas",
       x = "Grupo de Posição", y = "Contagem", fill = "Grupo de Posição") +
  guides(fill = guide_legend(title = "Grupo de Posição"))
print(hist_best_position2)

```



Agrupando as posições em grupos mais gerais.

```

players_by_country <- table(df$Nationality)
players_by_country_df <- as.data.frame(players_by_country)
names(players_by_country_df) <- c("country", "count")

world <- map_data("world")
world$count <- 0

players_by_country_df$country <- as.character(players_by_country_df$country)

country_mapping <- data.frame(
  original = c("Antigua & Barbuda", "Bosnia Herzegovina", "China PR", "Chinese Taipei",
              "Congo", "DR Congo", "England", "Guinea Bissau", "Hong Kong",

```

```

    "Korea DPR", "Korea Republic", "Macau", "Northern Ireland",
    "Republic of Ireland", "Saint Kitts and Nevis", "São Tomé & Príncipe",
    "Scotland", "Trinidad & Tobago", "United States", "Wales"),
corrected = c("Anguilla", "Bosnia and Herzegovina", "China", "China",
            "Republic of Congo", "Republic of Congo", "UK", "China",
            "China", "North Korea", "South Korea", "UK", "UK",
            "Ireland", "Nevis", "Sao Tome and Principe", "UK",
            "Trinidad and Tobago", "USA", "UK"))

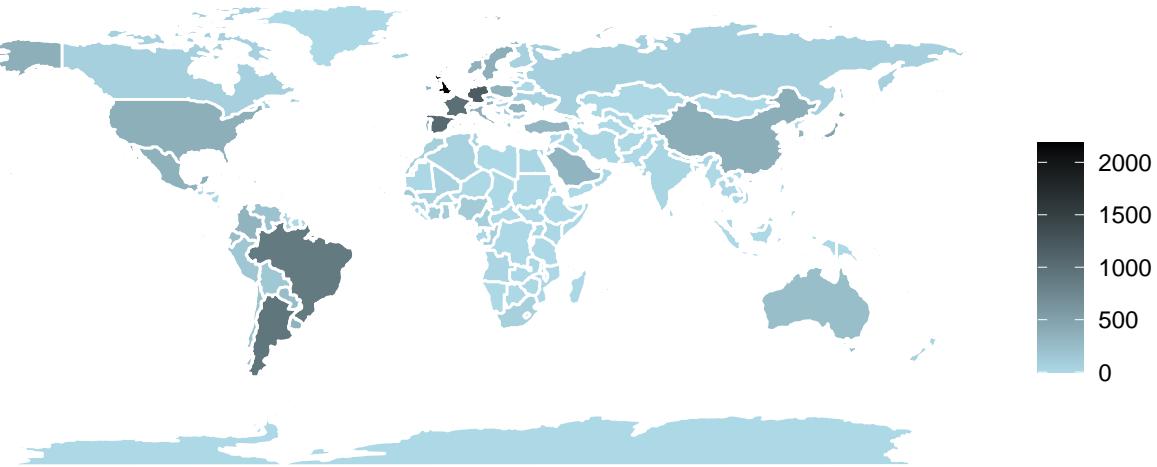
for (i in 1:nrow(country_mapping)) {
  players_by_country_df$country[players_by_country_df$country ==
    country_mapping$original[i]] <- country_mapping$corrected[i]

  players_by_country_df <- players_by_country_df %>%
    group_by(country) %>% summarize(count = sum(count))
  for (i in 1:nrow(world)) {
    country <- world$region[i]
    count <- players_by_country_df$count[players_by_country_df$country == country]
    if (length(count) > 0) {world$count[i] <- count)}

ggplot(world, aes(x = long, y = lat, group = group, fill = count)) +
  geom_polygon(color = "white") +
  scale_fill_gradient(low = "lightblue", high = "black") +
  theme_void() + coord_fixed() +
  labs(title = "Número de Jogadores por País", fill = "")

```

Número de Jogadores por País



O gráfico acima mostra o número de jogadores de futebol por país em um mapa mundi. Cada país é colorido de acordo com a quantidade de jogadores que representam aquele país no dataset. Países com mais jogadores aparecem em tons mais escuros, enquanto aqueles com menos jogadores têm tons mais claros.

Destaca-se que países como Brasil, Argentina e países da Europa como Reino Unido, Espanha, França e Alemanha são os mais numerosos em termos de jogadores de futebol no *dataset*. Isso é evidenciado pela intensidade mais escura de suas cores no mapa, indicando uma maior concentração de jogadores em relação a outros países. Esses países têm uma tradição consolidada no futebol e costumam ter uma presença significativa de jogadores em ligas de destaque ao redor do mundo.

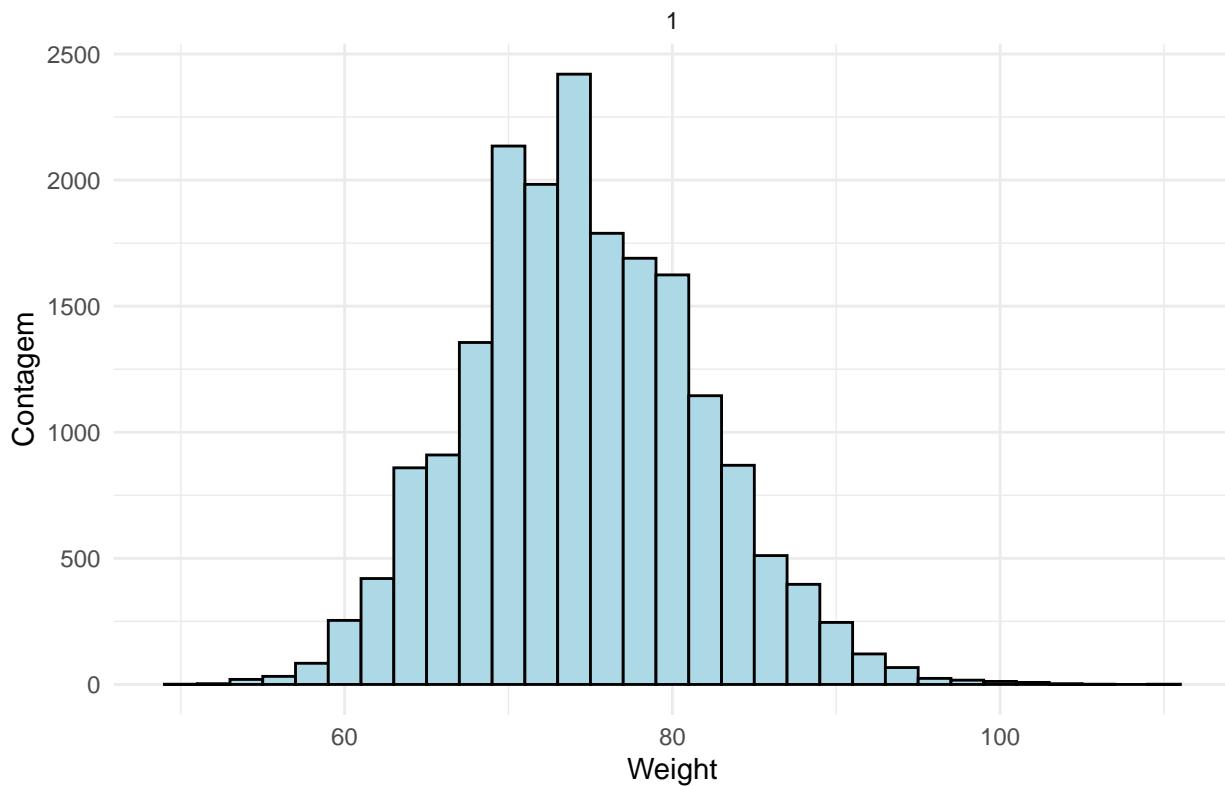
```

ggplot(df, aes(x = Weight)) +
  geom_histogram(binwidth = 2, fill = "lightblue", color = "black") +
  labs(title = "Distribuição de Peso", x = "Weight", y = "Contagem") +

```

```
facet_wrap(~1, ncol = 2) + theme_minimal()
```

Distribuição de Peso



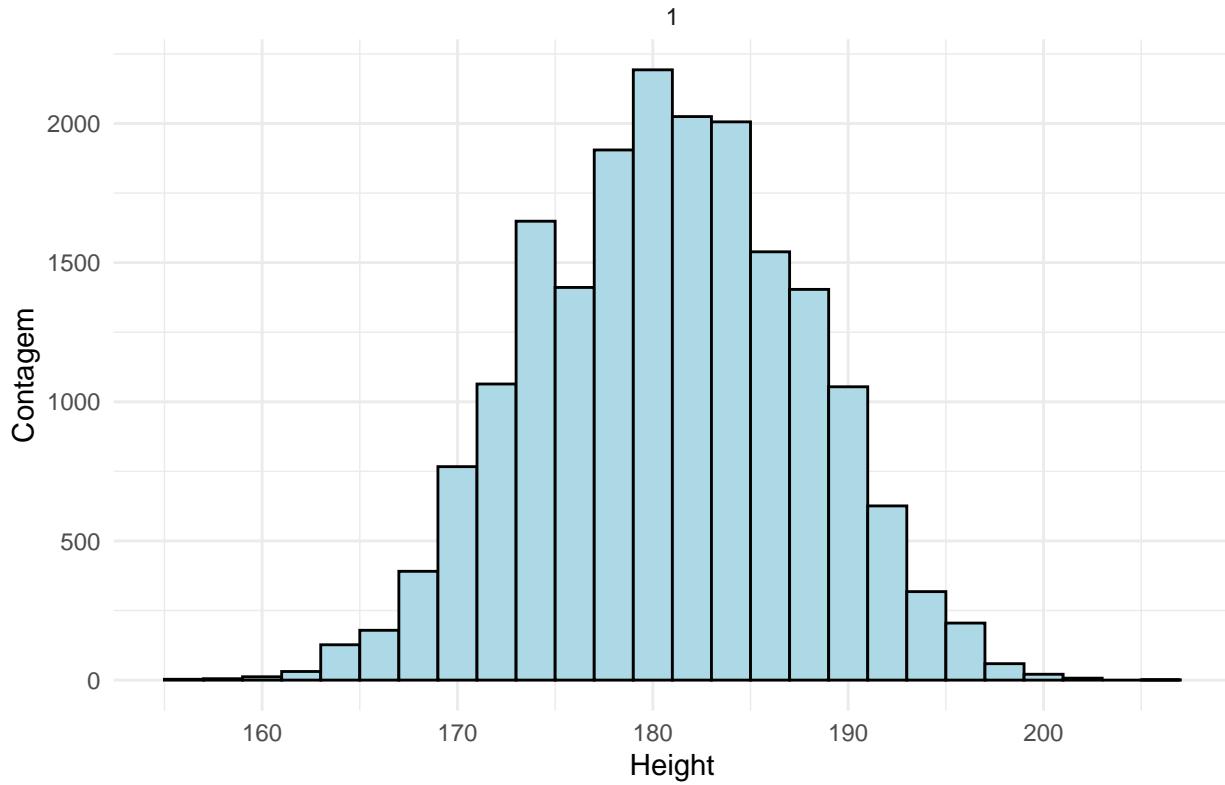
```
mean(df$Weight)
```

```
## [1] 75.04631
```

O gráfico acima mostra a distribuição de pesos dos jogadores no conjunto de dados. Observamos que a distribuição se assemelha a uma distribuição normal, com a maioria dos jogadores concentrados em torno de um peso médio de 75kg. Os pesos mais raros ou extremos estão nas extremidades do gráfico, enquanto a maior parte dos dados está centralizada em torno de um valor médio, sugerindo uma distribuição aproximadamente simétrica em torno da média.

```
ggplot(df, aes(x = Height)) +
  geom_histogram(binwidth = 2, fill = "lightblue", color = "black") +
  labs(title = "Distribuição de Altura", x = "Height", y = "Contagem") +
  facet_wrap(~1, ncol = 2) + theme_minimal()
```

Distribuição de Altura



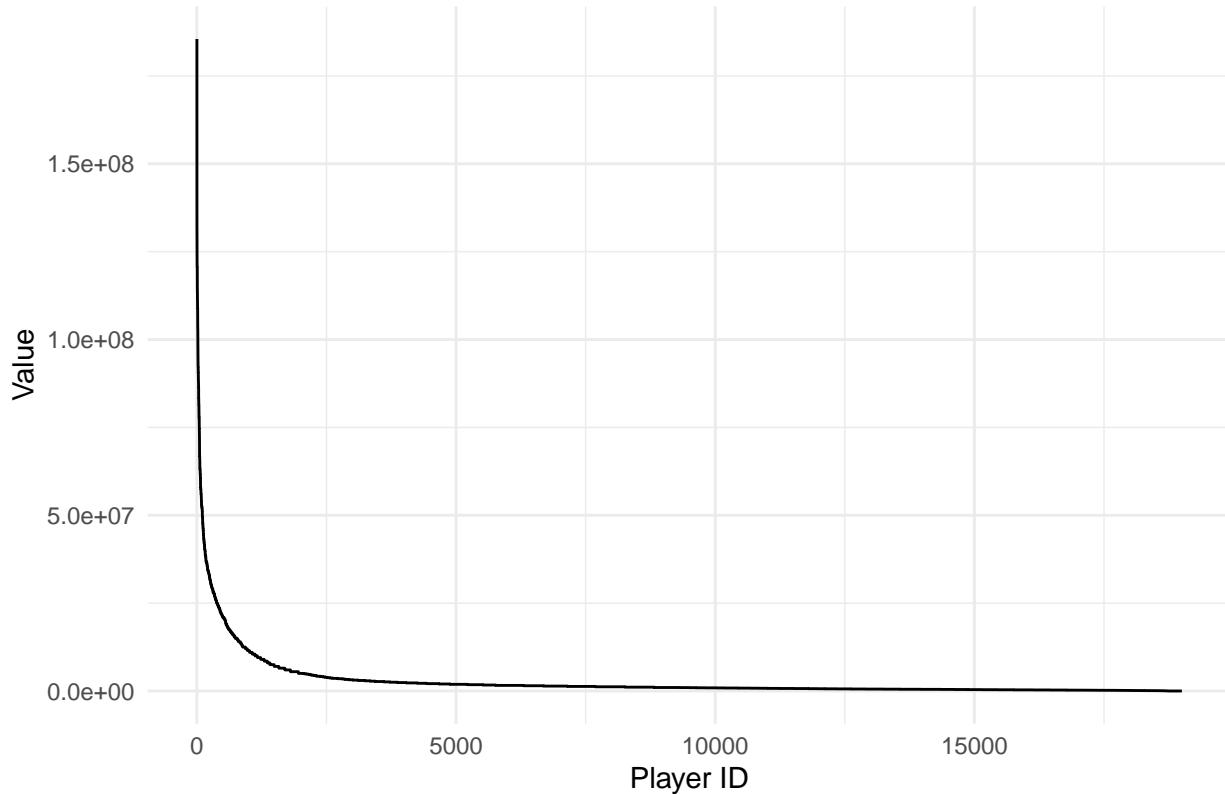
```
mean(df$Height)
```

```
## [1] 181.2257
```

O gráfico acima mostra a distribuição de altura dos jogadores no conjunto de dados. Observamos que a distribuição se assemelha a uma distribuição normal, com a maioria dos jogadores concentrados em torno de uma altura média de 1,81m.

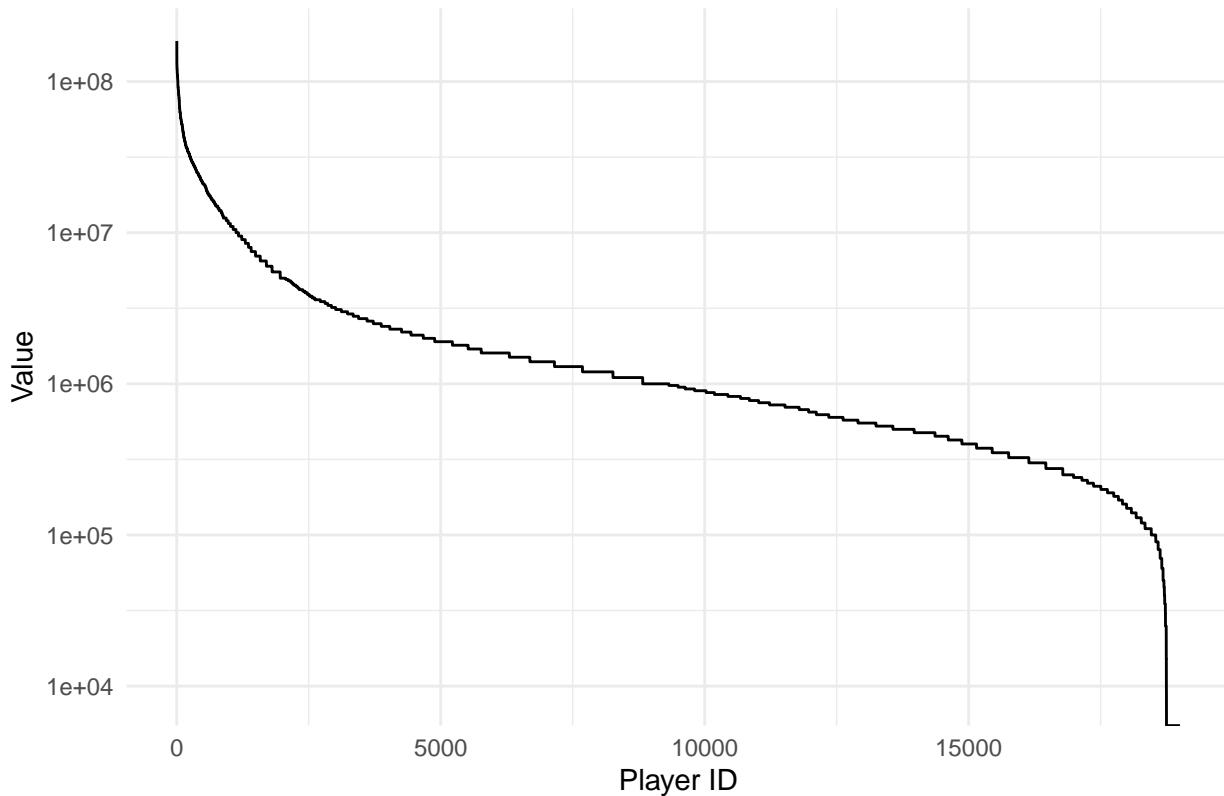
```
df_value <- df %>% arrange(desc(df$Value))
ggplot(df_value, aes(x = 1:nrow(df_value), y = Value)) +
  geom_line() +
  labs(title = "Gráfico de Valor", x = "Player ID", y = "Value") +
  theme_minimal()
```

Gráfico de Valor



```
ggplot(df_value, aes(x = 1:nrow(df_value), y = Value)) +  
  geom_line() +  
  labs(title = "Gráfico de Valor em Escala Logarítmica", x = "Player ID", y = "Value") +  
  scale_y_log10() + theme_minimal()
```

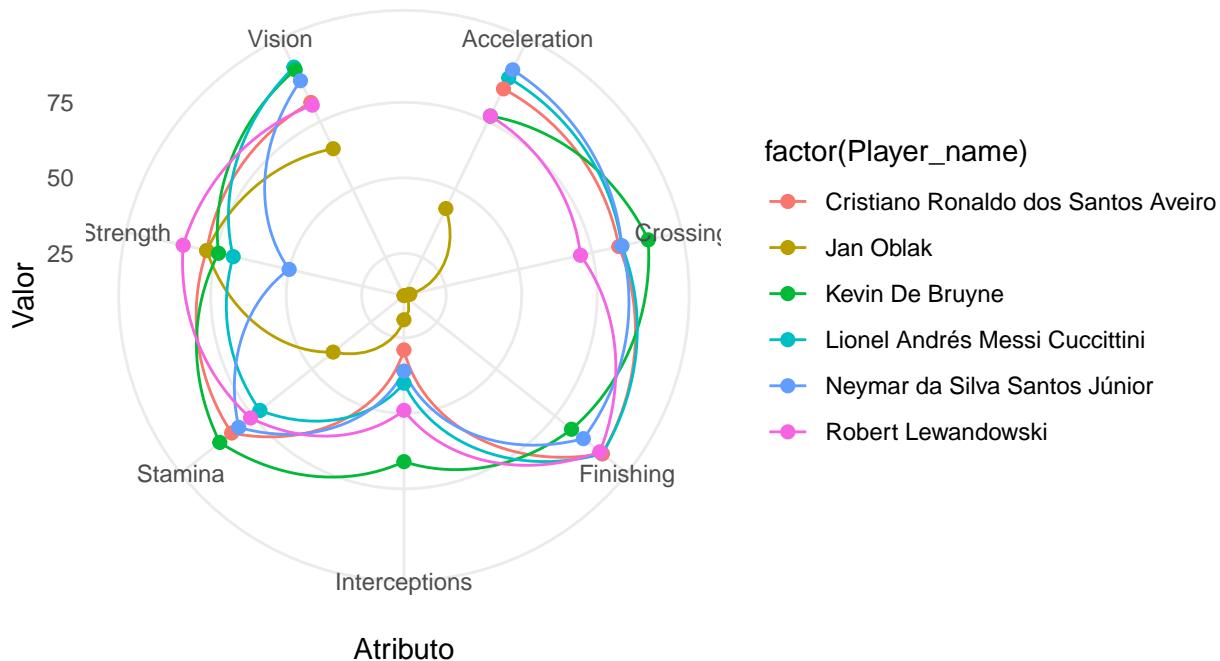
Gráfico de Valor em Escala Logarítmica



Nos gráficos acima vemos o comportamento da variável *Value*, também mostrada em escala logarítmica para melhor compreensão.

```
df_examples <- df[1:6, c("Player_name", "Finishing", "Crossing", "Interceptions", "Vision",  
                           "Acceleration", "Strength", "Stamina")]  
df_long <- df_examples %>%  
  pivot_longer(cols = -Player_name, names_to = "Attribute", values_to = "Value") %>%  
  group_by(Player_name)  
  
ggplot(df_long,  
       aes(x = Attribute, y = Value, group = Player_name, color = factor(Player_name))) +  
  geom_line() + geom_point(size = 2) +  
  labs(title = "Gráfico de Radar dos 10 Melhores Jogadores", x = "Atributo", y = "Valor") +  
  theme_minimal() + coord_polar()
```

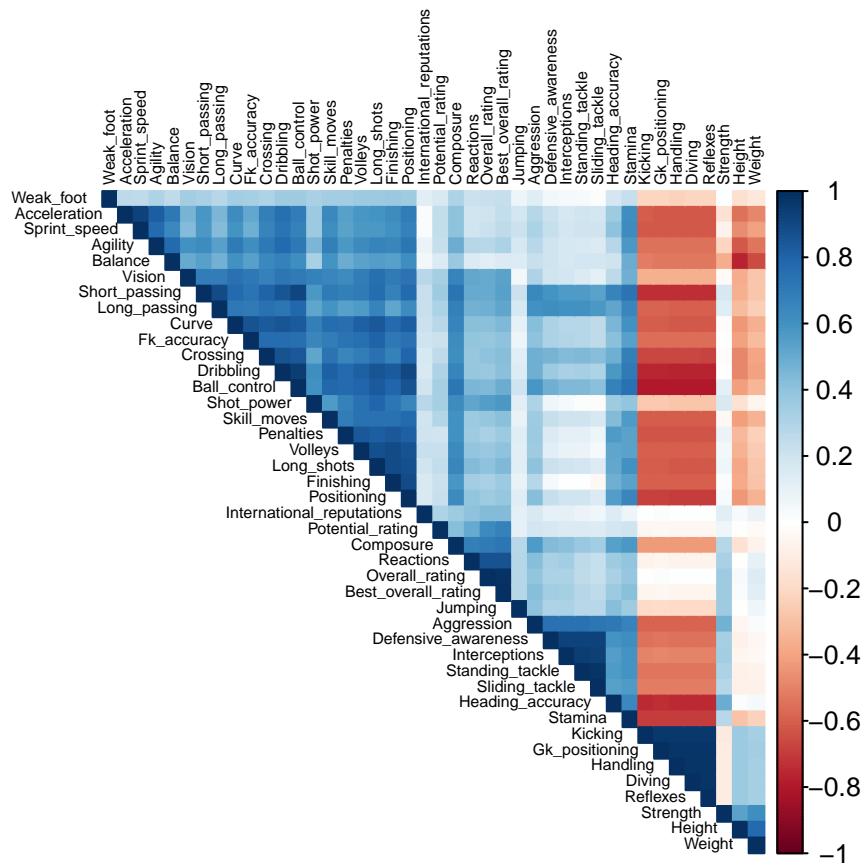
Gráfico de Radar dos 10 Melhores Jogadores



O gráfico acima é um gráfico de radar que representa os atributos dos seis melhores jogadores do *dataset*. Cada linha no gráfico corresponde a um jogador específico, e os diferentes atributos do jogador são representados nos eixos radiais. Os atributos incluem “Finishing” (Finalização), “Crossing” (Cruzamento), “Interceptions” (Interceptações), “Vision” (Visão), “Acceleration” (Aceleração), “Strength” (Força) e “Stamina” (Energia).

Este tipo de gráfico é útil para visualizar e comparar o desempenho de vários jogadores em diferentes atributos de uma maneira fácil de interpretar. Ele permite uma rápida compreensão de como cada jogador se destaca em diferentes áreas do jogo, como finalização, cruzamento, força, entre outros.

```
int_columns <- df %>% select(where(is.numeric)) %>%
    select(-c("Player_id", "Value", "Wage", "Team_id"))
corrplot(cor(int_columns), method = "color", type = "upper", tl.cex = 0.5,
        order = "hclust", tl.col = "black", tl.srt = 90, mar = c(0,0,0,0))
```



Este gráfico permite identificar padrões de correlação entre as variáveis numéricas do conjunto de dados. Correlações positivas são indicadas por cores mais claras, enquanto correlações negativas são indicadas por cores mais escuras. O *heatmap* facilita a identificação de quais variáveis estão mais fortemente correlacionadas entre si, o que pode fornecer insights interessantes para o desenvolvimento do *dashboard*.