Analisis NBA Players

Victor Benito Buendia

2026-01-13

```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE, fig.width=8, fig.height=5)
```

```
suppressPackageStartupMessages(library(tidyverse))
df <- read_csv("nba_globalizacion_2001_2026_FINAL.csv",
               col_types = cols(season = col_integer(), international = col_logical(),
                                mp = col_double(), pts = col_double(), trb = col_double(),
                                ast = col_double(), .default = col_guess()))
df <- df %>% mutate(international = as.logical(international),
                    net_rtg = o_rtg - d_rtg,
                    season = as.integer(season))
```

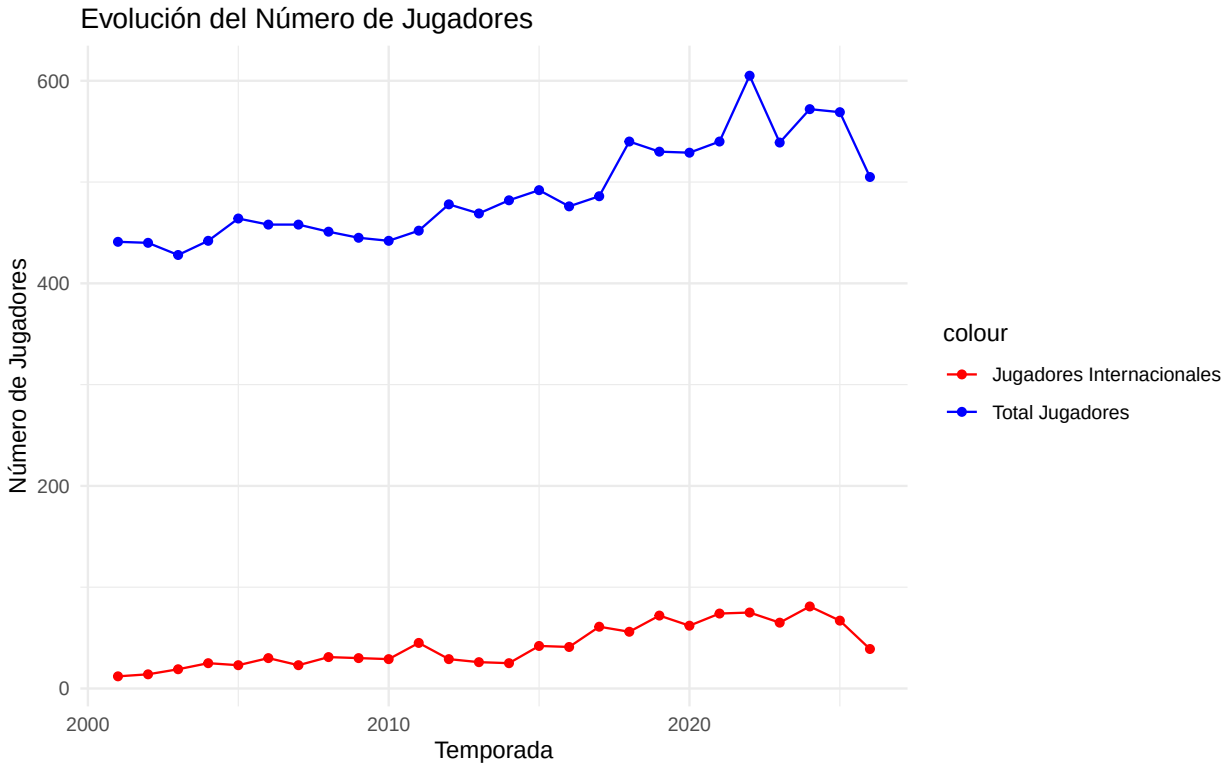
1. ¿Cómo ha evolucionado año tras año el número y el porcentaje de jugadores internacionales desde la temporada 2000-2001 hasta la actual 2025-2026?

```
q1 <- df %>%
  group_by(season) %>%
  summarise(total_players = n_distinct(player_id),
            intl_players = sum(international, na.rm = TRUE),
            pct_intl = (intl_players / total_players) * 100) %>%
  arrange(season)
print(q1)
```

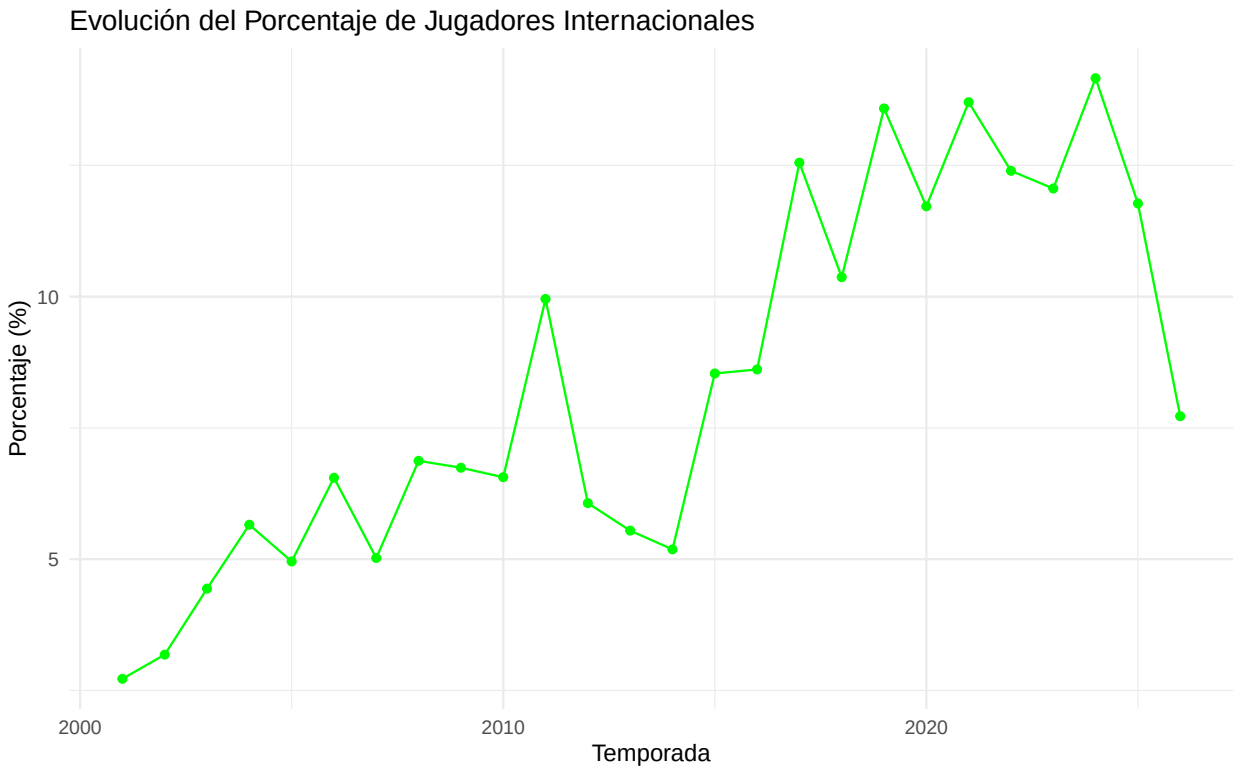
```
## # A tibble: 26 x 4
##    season total_players intl_players pct_intl
##     <int>         <int>        <int>    <dbl>
##  1   2001           441           12     2.72
##  2   2002           440           14     3.18
##  3   2003           428           19     4.44
##  4   2004           442           25     5.66
##  5   2005           464           23     4.96
##  6   2006           458           30     6.55
##  7   2007           458           23     5.02
##  8   2008           451           31     6.87
##  9   2009           445           30     6.74
## 10   2010           442           29     6.56
## # i 16 more rows
```

```
ggplot(q1, aes(x = season)) +
  geom_line(aes(y = total_players, color = "Total Jugadores")) +
  geom_point(aes(y = total_players, color = "Total Jugadores")) +
  geom_line(aes(y = intl_players, color = "Jugadores Internacionales")) +
  geom_point(aes(y = intl_players, color = "Jugadores Internacionales")) +
  labs(title = "Evolución del Número de Jugadores", x = "Temporada", y = "Número de Jugadores") +
  theme_minimal() +
```

```
  scale_color_manual(values = c("Total Jugadores" = "blue", "Jugadores Internacionales" = "red"))
```



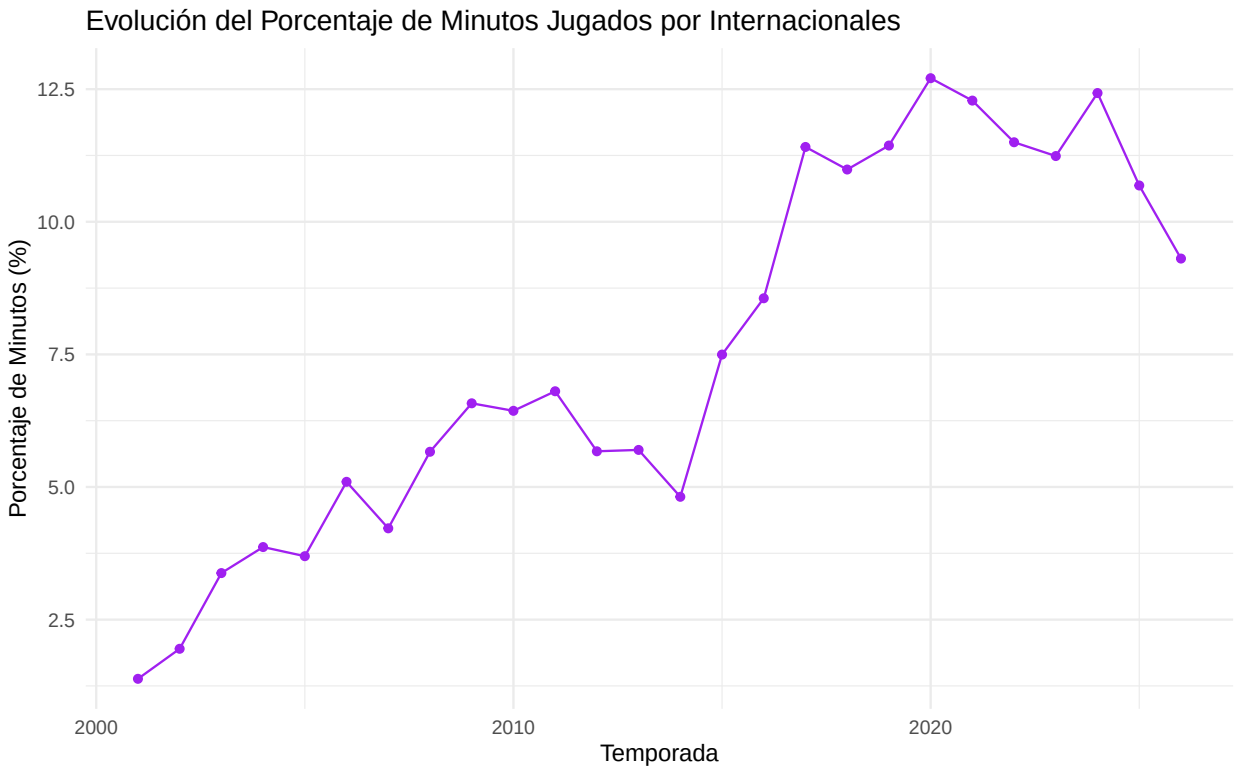

```
ggplot(q1, aes(x = season, y = pct_intl)) +
  geom_line(color = "green") +
  geom_point(color = "green") +
  labs(title = "Evolución del Porcentaje de Jugadores Internacionales", x = "Temporada", y = "Porcentaj
  theme_minimal()
```



2. ¿En qué medida ha aumentado el porcentaje de minutos jugados por internacionales en la liga?

```
q2 <- df %>%
  group_by(season) %>%
  summarise(total_mp = sum(mp, na.rm = TRUE),
            intl_mp = sum(intl_minutes, na.rm = TRUE),
            pct_intl_mp = (intl_mp / total_mp) * 100)
```

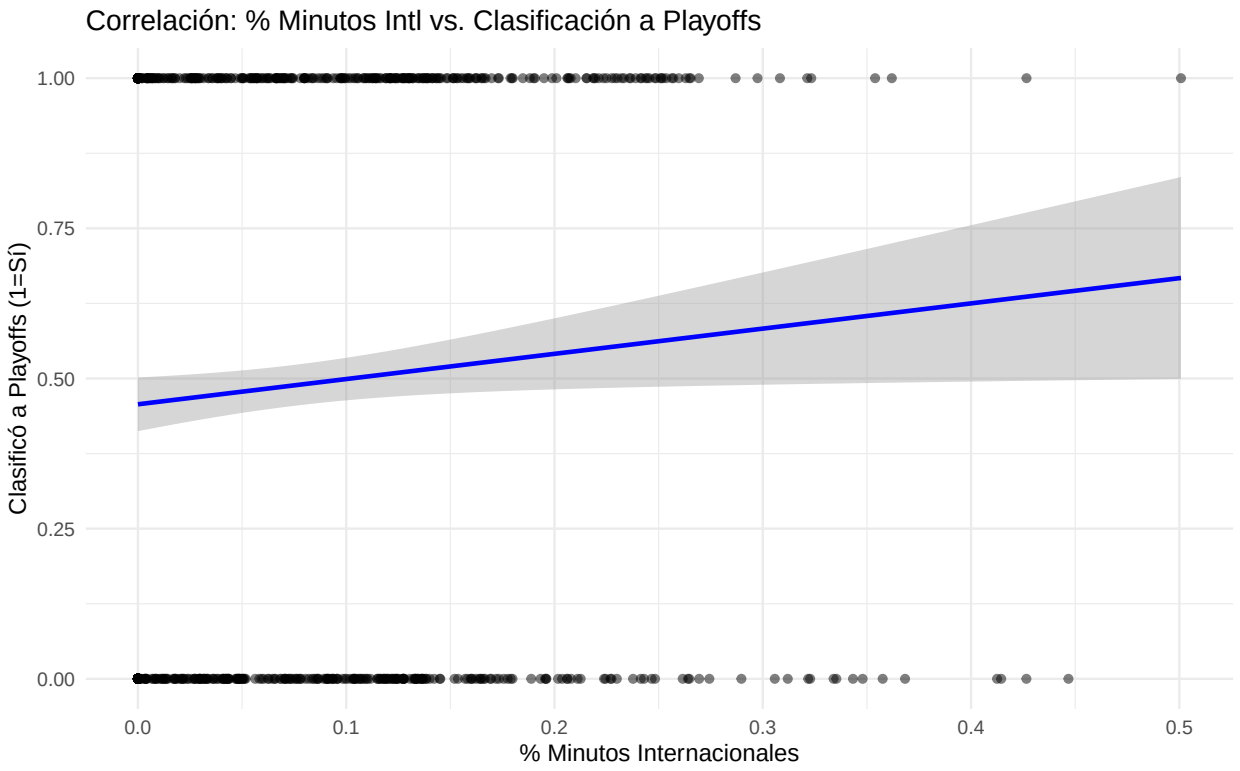
```
ggplot(q2, aes(x = season, y = pct_intl_mp)) +
  geom_line(color = "purple") +
  geom_point(color = "purple") +
  labs(title = "Evolución del Porcentaje de Minutos Jugados por Internacionales",
       x = "Temporada", y = "Porcentaje de Minutos (%)") +
  theme_minimal()
```



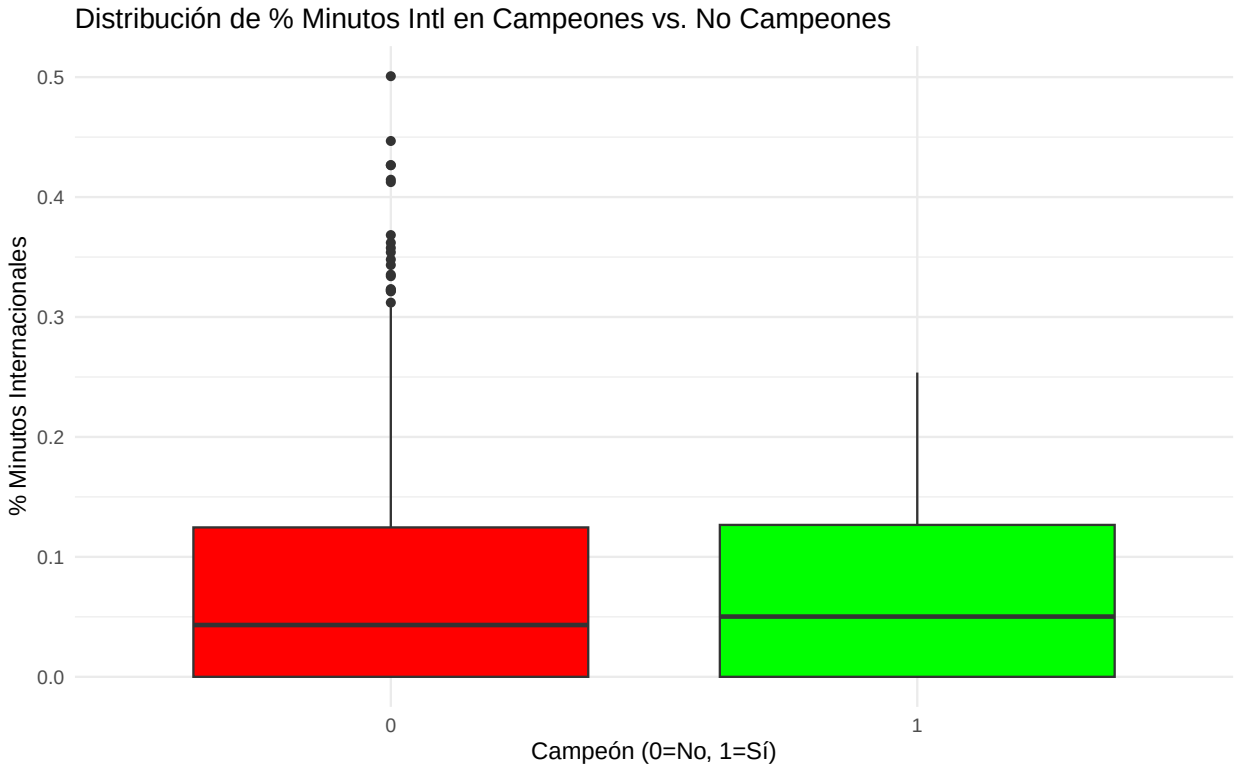
3. ¿Existe correlación entre el porcentaje de minutos internacionales en un equipo y su éxito (victorias en temporada regular y campeonatos)?

```
team_season <- df %>%
  group_by(season, team) %>%
  summarise(pct_intl_minutes = first(pct_intl_minutes_team),
            made_playoffs = max(in_playoffs),
            champion = max(champion),
            .groups = "drop")
```

```
ggplot(team_season, aes(x = pct_intl_minutes, y = made_playoffs)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", color = "blue") +
  labs(title = "Correlación: % Minutos Intl vs. Clasificación a Playoffs",
       x = "% Minutos Internacionales", y = "Clasificó a Playoffs (1=Sí)") +
  theme_minimal()
```




```
ggplot(team_season, aes(x = factor(champion), y = pct_intl_minutes)) +
  geom_boxplot(fill = c("red", "green")) +
  labs(title = "Distribución de % Minutos Intl en Campeones vs. No Campeones",
       x = "Campeón (0=No, 1=Sí)", y = "% Minutos Internacionales") +
  theme_minimal()
```



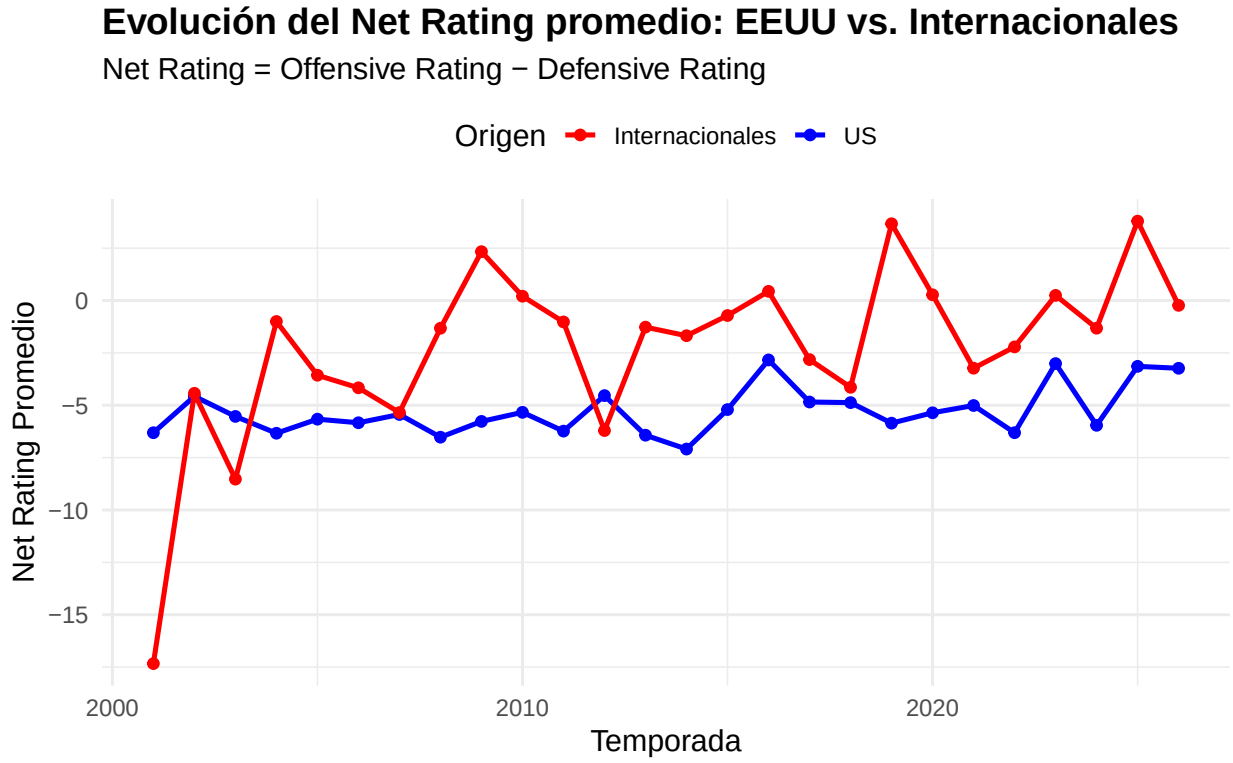
4. ¿En qué temporada los jugadores internacionales superaron por primera vez a los estadounidenses en indicadores avanzados de eficiencia (PER, WS/48, VORP)?

```
q4 <- df %>%
  group_by(season, international) %>%
  summarise(avg_net_rtg = mean(net_rtg, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = international, values_from = avg_net_rtg,
              names_prefix = "net_", values_fill = NA) %>%
  rename(net_US = net_FALSE, net_Intl = net_TRUE)
print(q4)
```

```
## # A tibble: 26 x 3
##    season net_US net_Intl
##     <int>  <dbl>    <dbl>
##  1   2001  -6.31  -17.3
##  2   2002  -4.57   -4.43
##  3   2003  -5.53   -8.53
##  4   2004  -6.34   -1
##  5   2005  -5.66   -3.57
##  6   2006  -5.83   -4.17
##  7   2007  -5.44   -5.35
##  8   2008  -6.52   -1.32
##  9   2009  -5.77    2.33
## 10   2010  -5.34    0.207
## # i 16 more rows
```

```
ggplot(q4, aes(x = season)) +
  geom_line(aes(y = net_US, color = "US"), linewidth = 1.1) +
  geom_point(aes(y = net_US, color = "US")) +
  geom_line(aes(y = net_Intl, color = "Internacionales"), linewidth = 1.1) +
```

```
  geom_point(aes(y = net_Intl, color = "Internacionales")) +
  labs(title = "Evolución del Net Rating promedio: EEUU vs. Internacionales",
       subtitle = "Net Rating = Offensive Rating - Defensive Rating",
       x = "Temporada", y = "Net Rating Promedio", color = "Origen") +
  scale_color_manual(values = c("US" = "blue", "Internacionales" = "red")) +
  theme_minimal(base_size = 14) +
  theme(plot.title = element_text(face = "bold"), legend.position = "top")
```



5. ¿Qué países y continentes han experimentado el mayor crecimiento en número de jugadores y calidad media durante este período?

```
intl_df <- df %>% filter(international, !is.na(continent), continent != "USA", continent != "")
cont_count <- intl_df %>%
  group_by(season, continent) %>%
  summarise(num_players = n_distinct(player_id), .groups = "drop") %>%
  complete(season = 2001:2026, continent = unique(continent), fill = list(num_players = 0))

growth_players <- cont_count %>%
  filter(season %in% c(2001, 2026)) %>%
  pivot_wider(names_from = season, values_from = num_players, names_prefix = "players_", values_fill = 
  mutate(growth = players_2026 - players_2001) %>%
  arrange(desc(growth))
print(growth_players)
```

```
## # A tibble: 6 x 4
##   continent     players_2001 players_2026 growth
##   <chr>                <int>        <int>  <int>
## 1 Europe                   4           16     12
## 2 North America            2           14     12
```

```
## 3 Africa                   1            6      5
## 4 Oceania                  0            2      2
## 5 Asia                     1            1      0
## 6 South America            0            0      0
```

```
cont_quality <- intl_df %>%
  group_by(season, continent) %>%
  summarise(avg_mp = mean(mp, na.rm = TRUE), .groups = "drop") %>%
  complete(season = 2001:2026, continent = unique(continent), fill = list(avg_mp = 0))

growth_quality <- cont_quality %>%
  filter(season %in% c(2001, 2026)) %>%
  pivot_wider(names_from = season, values_from = avg_mp, names_prefix = "mp_", values_fill = 0) %>%
  mutate(growth_mp = mp_2026 - mp_2001) %>%
  arrange(desc(growth_mp))
print(growth_quality)
```

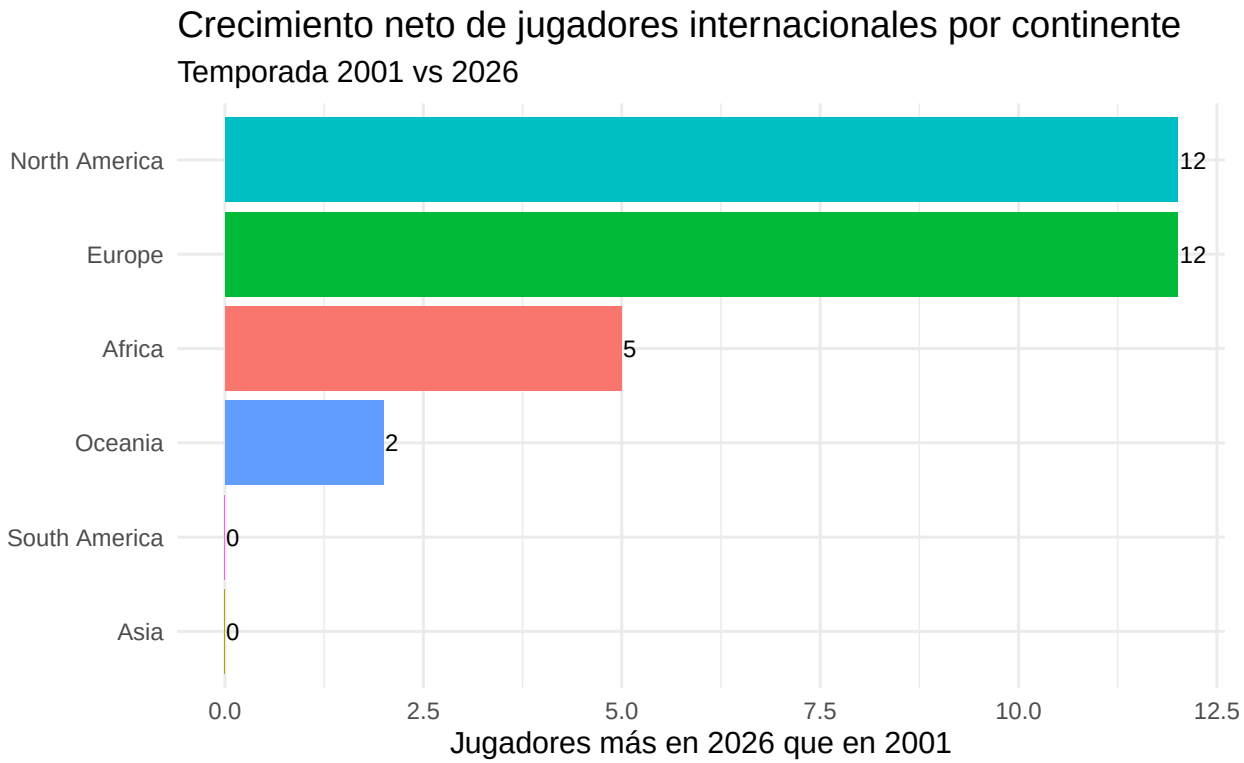
```
## # A tibble: 6 x 4
##   continent     mp_2001 mp_2026 growth_mp
##   <chr>           <dbl>   <dbl>     <dbl>
## 1 Oceania           0     562    562
## 2 Africa          221     548.   327.
## 3 Asia             38     161    123
## 4 North America   527     528.     0.571
## 5 South America     0       0      0
## 6 Europe         1302.    650.  -652.
```

```
country_growth <- intl_df %>%
  group_by(season, birth_country) %>%
  summarise(num_players = n_distinct(player_id), .groups = "drop") %>%
  complete(season = 2001:2026, birth_country = unique(birth_country), fill = list(num_players = 0)) %>%
  filter(season %in% c(2001, 2026)) %>%
  pivot_wider(names_from = season, values_from = num_players, names_prefix = "players_", values_fill =
  mutate(growth = players_2026 - players_2001) %>%
  arrange(desc(growth)) %>%
  filter(growth != 0)
print(country_growth)
```

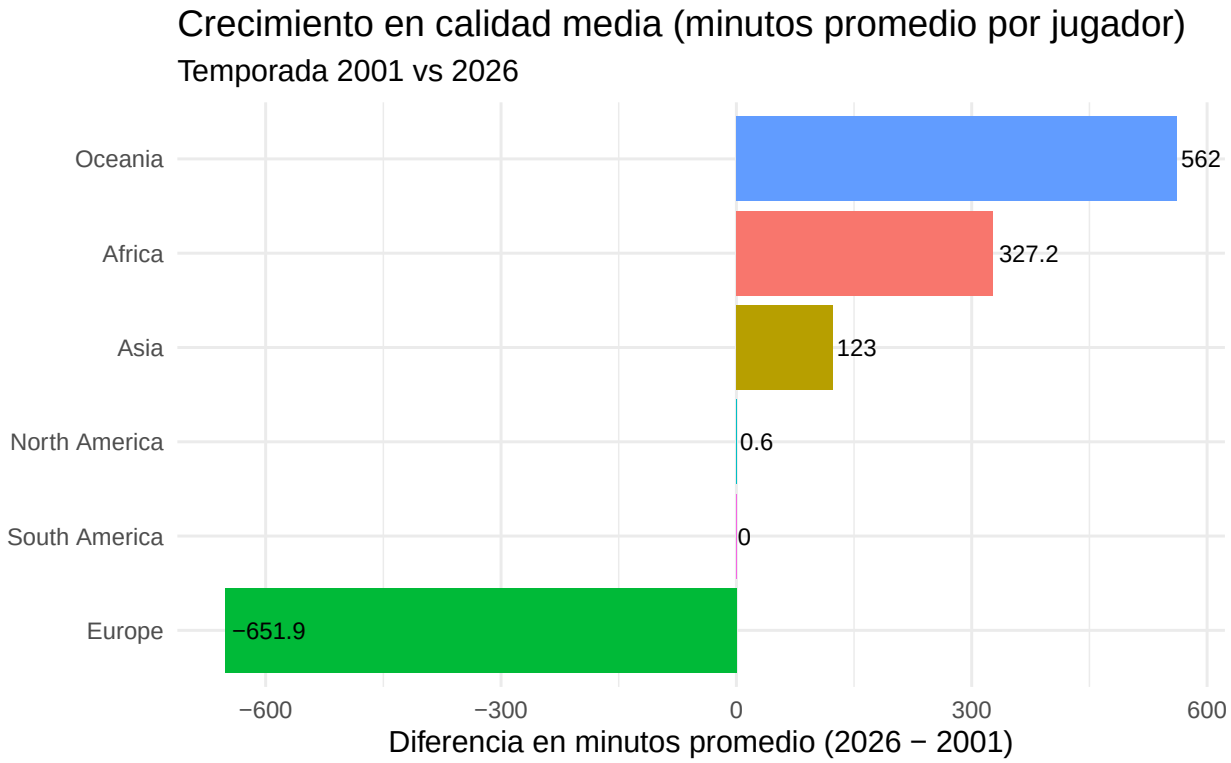
```
## # A tibble: 22 x 4
##    birth_country                    players_2001 players_2026 growth
##    <chr>                                   <int>        <int>  <int>
##  1 Canada                                      0           10     10
##  2 Australia                                   0            2      2
##  3 Bahamas                                     0            2      2
##  4 Democratic Republic of the Congo            0            2      2
##  5 France                                      0            2      2
##  6 Lithuania                                   0            2      2
##  7 Spain                                       0            2      2
##  8 Austria                                     0            1      1
##  9 Bosnia and Herzegovina                      0            1      1
## 10 Cameroon                                    0            1      1
## # i 12 more rows
```

```
ggplot(growth_players, aes(x = reorder(continent, growth), y = growth, fill = continent)) +
  geom_col() +
  geom_text(aes(label = growth), hjust = -0.1, size = 3.8) +
```

```
  coord_flip() +
  labs(title = "Crecimiento neto de jugadores internacionales por continente",
       subtitle = "Temporada 2001 vs 2026", x = NULL, y = "Jugadores más en 2026 que en 2001") +
  theme_minimal(base_size = 14) +
  theme(legend.position = "none")
```



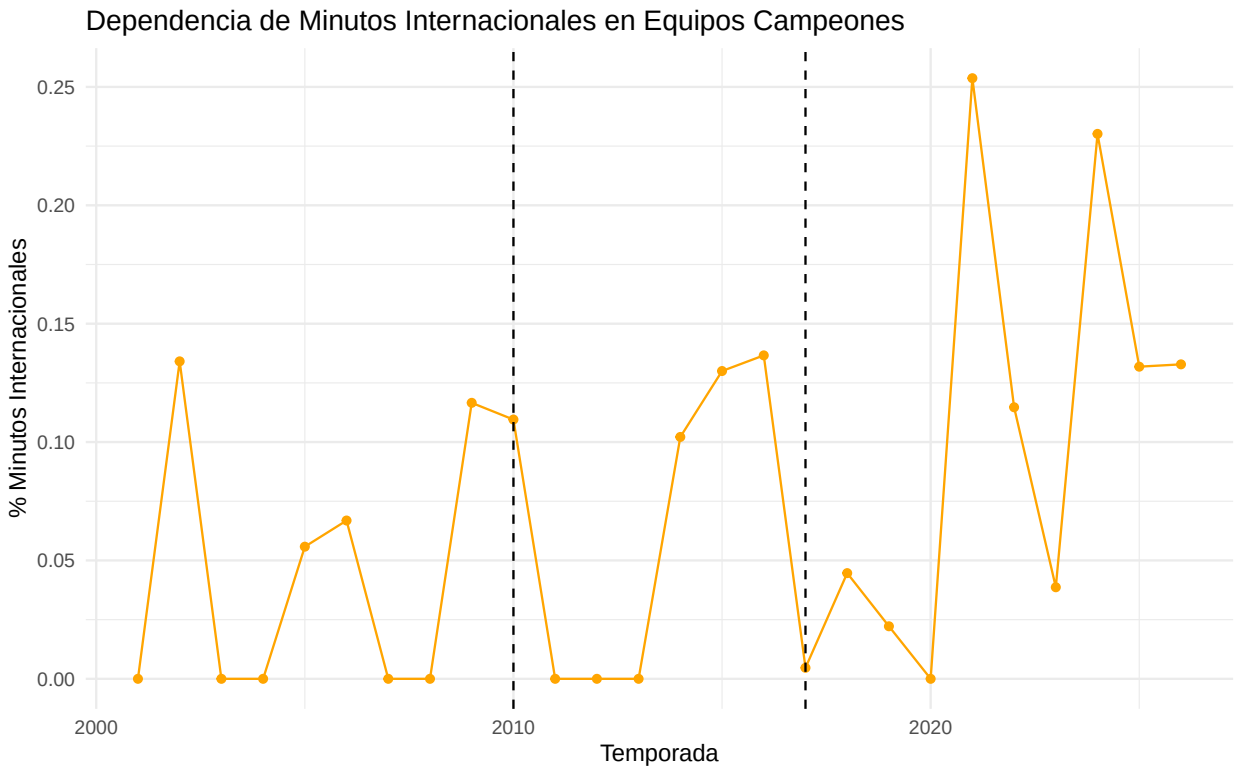

```
ggplot(growth_quality, aes(x = reorder(continent, growth_mp), y = growth_mp, fill = continent)) +
  geom_col() +
  geom_text(aes(label = round(growth_mp, 1)), hjust = -0.1, size = 3.8) +
  coord_flip() +
  labs(title = "Crecimiento en calidad media (minutos promedio por jugador)",
       subtitle = "Temporada 2001 vs 2026", x = NULL, y = "Diferencia en minutos promedio (2026 - 2001)")
  theme_minimal(base_size = 14) +
  theme(legend.position = "none")
```



6. ¿Los equipos campeones de la última década han dependido más del talento internacional que los campeones de principios del siglo XXI?

```
champ_teams <- df %>%
  filter(champion) %>%
  group_by(season) %>%
  summarise(team = first(team), pct_intl = first(pct_intl_minutes_team), .groups = "drop")
```

```
ggplot(champ_teams, aes(x = season, y = pct_intl)) +
  geom_line(color = "orange") +
  geom_point(color = "orange") +
  geom_vline(xintercept = c(2010, 2017), linetype = "dashed") +
  labs(title = "Dependencia de Minutos Internacionales en Equipos Campeones",
       x = "Temporada", y = "% Minutos Internacionales") +
  theme_minimal()
```


Dependencia de Minutos Internacionales en Equipos Campeones
% Minutos Internacionales
0.25
0.20
0.15
0.10
0.05
0.00
2000
2010
2020
Temporada