International Matchups EDA

Ben Wieland

2024-02-05
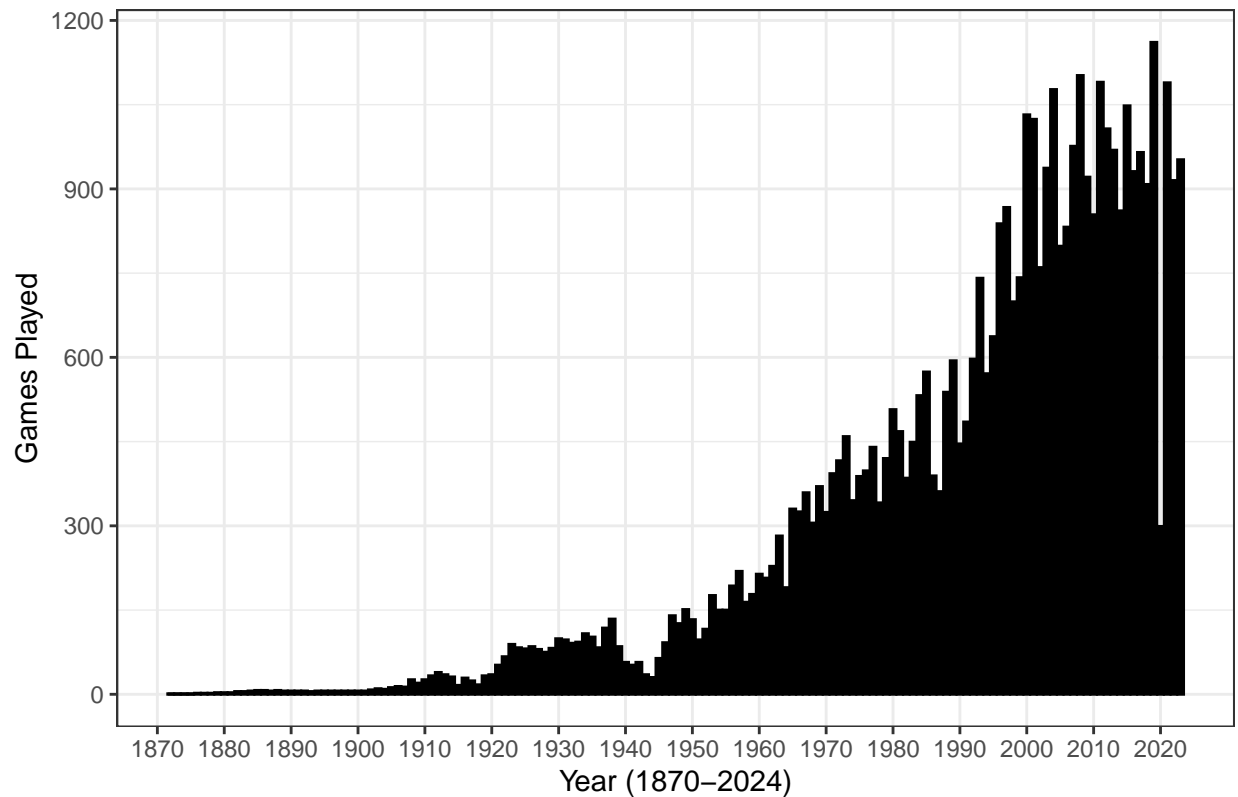
```r
source("load-data.R")
```

```r
theme_plots <- function() {
  theme_bw()
}

theme_set(theme_plots())
```

```r
ggplot(results, aes(x = year(date))) +
  geom_bar(color = "black", fill = "black") +
  scale_x_continuous(breaks = seq(1870, 2020, 10)) +
  theme(panel.grid.minor.x = element_blank()) +
  labs(x = "Year (1870-2024)",
       title = "International Soccer Matches by Year",
       y = "Games Played")
```

International Soccer Matches by Year


```r
results %>%
  group_by(tournament) %>%
  summarize(matches = n(),
            first_played = min(year(date)),
            last_played = max(year(date))) %>%
  arrange(-matches) %>%
  mutate(matches = formatC(matches, format = "d", big.mark = ",")) %>%
  head(10) %>%
  as.data.frame()
```

```
##                             tournament matches first_played last_played
## 1                             Friendly  17,773         1872        2023
## 2           FIFA World Cup qualification   8,016         1933        2023
## 3              UEFA Euro qualification   2,815         1958        2023
## 4  African Cup of Nations qualification   1,998         1961        2023
## 5                        FIFA World Cup     964         1930        2022
## 6                          Copa América     841         1916        2021
## 7           AFC Asian Cup qualification     764         1956        2022
## 8               African Cup of Nations     741         1957        2022
## 9                            CECAFA Cup     620         1973        2017
## 10        CFU Caribbean Cup qualification     606         1978        2017
```

```r
home_results <- results %>%
  group_by(year = year(date)) %>%
  summarise(Win = length(home_result[home_result == "W"]),
```

```
                Draw = length(home_result[home_result == "D"]),
                Loss = length(home_result[home_result == "L"])) %>%
    pivot_longer(cols = c(Win, Draw, Loss), names_to = "result",
                 values_to = "count") %>%
    mutate(result = factor(result, levels = c("Win","Draw","Loss")))

ggplot(home_results %>% filter(year >= 1960), aes(x = year, y = count,
                                                  fill = result),
       color = "white") +
  geom_col(position = "fill") +
  scale_fill_manual(values = c("Win" = "darkgreen",
                               "Draw" = "grey",
                               "Loss" = "tomato")) +
  coord_flip() +
  scale_x_reverse() +
  labs(title = "Home Field Advantage Remains Nearly Constant over Time",
       y = "Proportion of Results for Home Team",
       x = "Year (1960-2024)") +
  scale_y_continuous(labels = scales::percent)
```


Home Field Advantage Remains Nearly Constant over Time