

Informe práctico 2

Noelia Avalos

2023-09-13

Avalos Noelia.

La base de datos `resultados_cmff` contiene información sobre las Copas Mundiales de Fútbol Femenino y su fuente es la `data.world`. En ese sitio encontrarás más datos en inglés si te interesa profundizar.

Parte 1

Pregunta:

¿Cómo es la distribución geográfica de los goles por países? ¿Y la distribución de goles a través del tiempo?

Introducción:

En el presente informe se tomará el conjunto de datos “`resultados_cmff`” proveniente de `data.world`, el cual contiene información sobre las Copas Mundiales de Fútbol Femenino. Esta base de datos proporciona detalles sobre los equipos, rondas, resultados y los goles anotados en cada torneo. Nuestro objetivo es analizar la distribución geográfica de goles y la distribución de goles a través del tiempo. Para responder el primer interrogante utilizaremos las columnas “`codigo_pais`” para identificar los países y “`goles`” para calcular la cantidad de goles por país.

Por otro lado, para analizar la distribución de goles a lo largo del tiempo, nos basaremos en la columna “`anio`” para identificar el año de cada Copa Mundial y en la columna “`goles`” para contar el número total de goles anotados en ese año.*

Enfoque:

Para encontrar la distribución geográfica de goles por país, realizaremos un choropleth map del planisferio. Dicho mapa mostrará de manera visual e intuitiva la distribución geográfica de los goles por países, donde los países con más goles tendrán un color más intenso, mientras que los países con menos goles tendrán un color más claro. El uso de colores mostrará una representación visual inmediata de la cantidad de goles anotados por cada país. De esta manera podremos identificar de forma rápida y efectiva los países con mayor cantidad de anotaciones.

Luego, utilizaremos una tabla para analizar la distribución de goles a lo largo del tiempo con el fin de tener una visión más cuantitativa y desglosada de los datos. Al agrupar los goles por año y sumarlos, obtendremos una visión general del número total de goles anotados cada año. De esta forma, podremos analizar la evolución de los goles a lo largo de la competencia y detectar posibles tendencias.

Análisis:

```
ggthemr("earth")
url <- "https://simple.wikipedia.org/wiki/Comparison_of_IOC,_FIFA,_and_ISO_3166_country_codes"
lista_fifa_iso <- url %>%
  read_html() %>%
  html_table(fill = TRUE) %>% .[[1]] %>%
```

```

mutate(ISO = case_when(Country == "England" ~ "ENG",
                        Country == "Scotland" ~ "SCT",
                        TRUE ~ ISO)) %>%

select(FIFA, ISO)

resultados_combinados <- resultados_cmff %>%
  left_join(lista_fifa_iso, by = c("codigo_pais" = "FIFA"))

#Se realiza el calculo para conocer la distribucion de goles
goles_paises <- resultados_combinados %>%
  group_by(ISO) %>%
  summarise(cant_goles = sum(goles)) %>%
  arrange(desc(cant_goles))

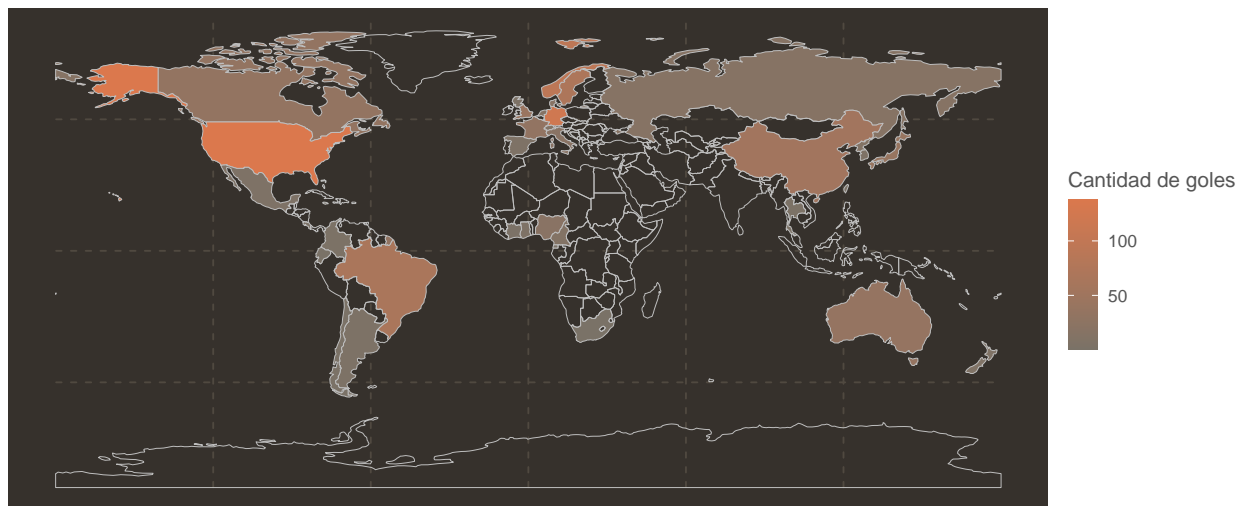
#Llamamos a los mapas necesarios para visualizar
mapa_mundi <- rnatrualearth::ne_countries(returnclass = "sf") %>%
  select(geometry,gu_a3)
mapa_uk <- rnatrualearth::ne_countries(country = "united kingdom", type = "map_units", returnclass = 'sf') %>%
  select(geometry,gu_a3)

mapa_combinado <- rbind(mapa_mundi, mapa_uk)
mapa_goles <- left_join(mapa_combinado, goles_paises, by = c("gu_a3" = "ISO"))

ggplot() +
  geom_sf(data = mapa_goles, aes(fill = cant_goles), color = "gray") +
  labs(title = "Distribución geográfica de anotaciones por equipos",
       fill = "Cantidad de goles",
       caption = "Fuente: data.world")

```

Distribución geográfica de anotaciones por equipos



Fuente: data.world

```
goles_por_anio <- resultados_cmff %>%
  group_by(anio) %>%
  summarise(total_goles = sum(goles))

tabla_goles_por_anio <- goles_por_anio %>%
  kable(col.names = c("Año", "Total de goles"),
        caption = "Total de goles por año",
        align = "r") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive")) %>%
  row_spec(c(3,5,7,8), background = "#C78A59") %>%
  column_spec(1, width = "50px")

tabla_goles_por_anio
```

Discusión:

Visualizando el mapa, podemos ver una gran concentración de color y por lo tanto, de goles en Estados Unidos, seguido de Alemania, Noruega y Suecia. Sin embargo, en la distribución geográfica el color va decreciendo en países como Brasil, Inglaterra, China y Japón, finalmente el color carece de intensidad en los países con menor cantidad de goles.

Al analizar la tabla de distribución de goles a través del tiempo, identificamos los años con un mayor registro de goles: 1999, 2007, 2015 y 2019, con valores de 123, 111, 148 y 145 respectivamente, los años restantes tuvieron

Table 1: Total de goles por año

Año	Total de goles
1991	98
1995	99
1999	123
2003	107
2007	111
2011	86
2015	148
2019	145

una menor cantidad de goles anotados. Sin embargo, estas fluctuaciones entre años, no sugieren una tendencia de aumento o disminución sostenida.

Parte 2

Pregunta: ¿Cuál es el índice de victorias a nivel continental? ¿Tiene relación con el desempeño por equipo a través del tiempo?

Introducción:

El objetivo de esta parte es identificar los continentes con mayores victorias y analizar si hay equipos de esos continentes que hayan mejorado su desempeño a lo largo del tiempo. Para esto, utilizaremos la columna “resultados” agrupando por “continente” el número de victorias que tiene dicho equipo.

En segunda instancia y teniendo en cuenta que el análisis de desempeño de un equipo no siempre se refleja en el resultado de los partidos, como por ejemplo el caso de Brasil en el Mundial Masculino 2022, donde anotaron más goles que Argentina pero no lograron pasar la fase de cuartos de final. Entonces considerando que el desempeño por equipo, puede depender de otras variables, como la estrategia, niveles ofensivos y defensivos, etc. Utilizaremos la columna “goles” para medir dicho desempeño ya que es un “indicador de éxito” del equipo, dividiendo el periodo de tiempo en dos, el primer periodo se dará hasta el año 2007 y el segundo periodo desde 2007

Enfoque:

Abordando la primera parte de la pregunta, utilizaremos un gráfico de barras para analizar el desempeño a nivel continental en base a los resultados ya que buscamos comparar de manera clara y directa la cantidad de victorias por continente.

En la segunda parte, realizaremos un gráfico de dispersión donde calcularemos el desempeño por equipo en base al promedio de goles hasta el año 2007, luego repetiremos el proceso con un segundo gráfico de dispersión desde el año 2007. En última instancia los combinaremos con el fin de comparar el rendimiento por equipos y su evolución a lo largo del tiempo.

Análisis:

```
url2 <- "https://simple.wikipedia.org/wiki/Comparison_of_IOC,_FIFA,_and_ISO_3166_country_codes"
lista_pais <- read_html(url2) %>%
  html_table(fill = TRUE) %>% .[[1]] %>%
  select(FIFA,Country)

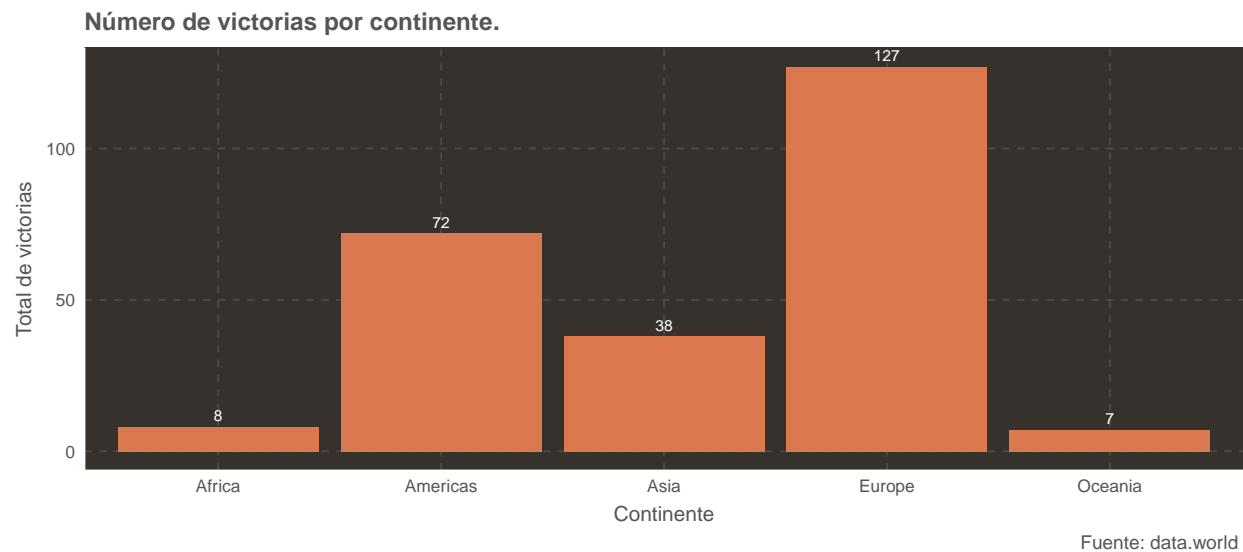
resultados_paises <- resultados_cmff %>%
  left_join(lista_pais, by = c("codigo_pais" = "FIFA"))
resultados_cmff_con_continentes <- resultados_paises %>%
```

```

mutate(continente = ifelse(equipo %in% c("Inglaterra", "Escocia"), "Europe", countrycode(sourcevar = 
))
# Filtrar y contar las victorias por continente
victorias_por_continente <- resultados_cmff_con_continentes %>%
  filter(resultado == "Victoria") %>%
  group_by(continente) %>%
  summarise(total_victorias = n())

# Visualizar los resultados
ggplot(victorias_por_continente, aes(x = continente, y = total_victorias)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = total_victorias), vjust = -0.5, color = "white", size = 3) +
  labs(title = "Número de victorias por continente.", x = "Continente", y = "Total de victorias",
        caption = "Fuente: data.world")

```



```

#Realizamos operaciones para encontrar el desempeño basado en anotaciones hasta el año 2007
resultados_hasta_2007 <- resultados_combinados %>%
  filter(anio <= 2007) %>%
  group_by(equipo) %>%
  summarise(desempeno = mean(goles)) %>%
  arrange(desc(desempeno))

#Filtramos el equipo con mejor desempeño
mejor_equipo_1 <- resultados_hasta_2007 %>%
  filter(desempeno == max(desempeno))

#Armamos el primer grafico
grafico1 <- ggplot(resultados_hasta_2007, aes(desempeno, equipo)) +
  geom_jitter(position = "identity") +
  geom_text(data = mejor_equipo_1, aes(label = paste0(round(desempeno, 2))), nudge_y = 1.3, size = 3) +
  geom_point(data = mejor_equipo_1, shape = 21, color = "white", fill = NA, aes(color = "Mejor Equipo"),
  labs(title = "Desempeño hasta 2007",
        x = "Desempeño",
        y = "Equipo") +

```

```

theme(plot.title = element_text(size = 10)) +
coord_cartesian(xlim = c(0, 3))

#Comenzamos a analizar desde 2007
resultados_desde_2007 <- resultados_combinados %>%
  filter(anio >= 2007) %>%
  group_by(equipo) %>%
  summarise(desempeno = mean(goles)) %>%
  arrange(desc(desempeno))

#Filtramos el equipo con mejor desempeño
mejor_equipo_2 <- resultados_desde_2007 %>%
  filter(desempeno == max(desempeno))

#Armamos el segundo grafico
grafico2 <- ggplot(resultados_desde_2007, aes(desempeno, equipo)) +
  geom_jitter(position = "identity") +
  geom_text(data = mejor_equipo_2, aes(label = paste0(desempeno)), nudge_y = 1.3, size = 3) +
  geom_point(data = mejor_equipo_2, shape = 21, color = "white", fill = NA, aes(color = "Mejor Equipo"),
  labs(title = "Desempeño desde 2007",
        x = "Desempeño",
        y = "Equipo") +
  theme(plot.title = element_text(size = 10)) +
  coord_cartesian(xlim = c(0, 3))

# Combinar los gráficos y agregar títulos
(grafico1 | grafico2) +
  plot_annotation(title = "Comparativa del desempeño en anotaciones por equipo.",
                  caption = "Fuente: data.world")

```

Comparativa del desempeño en anotaciones por equipo.



““

Discusión:

En el primer grafico podemos ver que el indice de victorias a nivel continental, es superior en Europa y las Americas, siendo Europa, superior con un total de 127 victorias. Aqui vemos que el nivel continental,

podemos relacionarlo tambien con la distribucion geografica de goles, si bien Estados Unidos era el equipo con mayor cantidad de goles, los que lo seguian en terminos de cantidad eran paises Europeos.

Cuando analizamos el desempeño por equipo en dos periodos de tiempos determinados, vemos que antes de 2007, el pais con mejor desempeño es Alemania, con un promedio de 2.96 de anotaciones, despues de 2007, vemos que mas paises se suman a la competencia, sin embargo el pais predominante sigue siendo Europeo, en este caso Suiza con un promedio de 2.75. Tambien existen paises que mejoraron su desempeño a lo largo del tiempo como Estados Unidos, Brasil, Italia y otros que descendieron, al igual que Alemania, Noruega y la Republica popular China. Estos resultados sugieren que el desempeño continental está relacionado con el desempeño por equipo a lo largo del tiempo. Los continentes que han demostrado una gran cantidad de victorias a nivel continental también tienen equipos destacados en términos de promedio de goles por partido.