

Segunda Entrega del Proyecto

28/3/2021

Buen desempeño económico, la clave del éxito para un rendimiento sobresaliente en los juegos olímpicos.

1. Integrantes de grupo.

- *Nicolas González*
- *Joan Galeano*
- *Alejandro Guevara*

Paquetes

```
library(tidyverse)
library(rvest)
library(haven)
library(wbstats)
library(dplyr)
library(naniar)
library(knitr)
library(ggthemes)
library(readxl)
library(GGally)
library(Hmisc)
library(corrplot)
library(PerformanceAnalytics)
```

Manejo de la base de datos

Variable explicada

```
pagina <- "http://www.olympedia.org/statistics/medal/country"
pagina_desc <- read_html(pagina)

países <- pagina_desc %>% html_nodes("td:nth-child(1)") %>% html_text()

medallas <- pagina_desc %>% html_nodes("td:nth-child(6)") %>% html_text()
medallas <- as.integer(medallas)
medallas_por_pais <- tibble(países, medallas)
```

```

medallas_por_pais[4,1] <- "United Kingdom"
medallas_por_pais[7,1] <- "China"
medallas_por_pais[18,1] <- "Korea, Rep."
medallas_por_pais[43,1] <- "Iran, Islamic Rep."
medallas_por_pais[50,1] <- "Slovak Republic"
medallas_por_pais[59,1] <- "Egypt, Arab Rep."
medallas_por_pais[62,1] <- "Bahamas, The"
medallas_por_pais[83,1] <- "Venezuela, RB"
medallas_por_pais[84,1] <- "Serbia"
medallas_por_pais[97,1] <- "Cote d'Ivoire"
medallas_por_pais[98,1] <- "Hong Kong SAR, China"
medallas_por_pais[113,1] <- "Moldova"
medallas_por_pais[117,1] <- "Tanzania"
medallas_por_pais[121,1] <- "Kyrgyz Republic"
medallas_por_pais[122,1] <- "Saudi Arabia"

```

Variables explicativas

En este caso, se presenta la Tasa de crecimiento del PIB como primera variable explicativa.

```

growth_gdp <- wb_data("NY.GDP.PCAP.KD.ZG", start_date = 1950, end_date = 2016)
growth_gdp <- tibble(growth_gdp$country, growth_gdp$date, growth_gdp$NY.GDP.PCAP.KD.ZG)
growth_gdp <- growth_gdp %>%
  rename(
    paises = `growth_gdp$country`,
    fecha = `growth_gdp$date`,
    growth = `growth_gdp$NY.GDP.PCAP.KD.ZG`
  )
growth_gdp <- drop_na(growth_gdp)

growth_gdp_prom <- aggregate(growth_gdp$growth, list(growth_gdp$paises), FUN=mean)
growth_gdp_prom <- growth_gdp_prom %>%
  rename(
    paises = Group.1, GDP=x
  )

```

A continuación, se usará la Tasa de crecimiento de la población.

```

growth_pob <- wb_data("SP.POP.GROW", start_date = 1950, end_date = 2016)
growth_pob <- tibble(growth_pob$country, growth_pob$date, growth_pob$SP.POP.GROW)
growth_pob <- growth_pob %>%
  rename(
    paises = "growth_pob$country",
    fecha = "growth_pob$date",
    growth_p = "growth_pob$SP.POP.GROW"
  )
growth_pob <- drop_na(growth_pob)

growth_pob_prom <- aggregate(growth_pob$growth_p, list(growth_pob$paises), FUN=mean)
growth_pob_prom <- growth_pob_prom %>%
  rename(

```

```
países = Group.1, POB=x
)
```

También se decidió incluir la Tasa promedio de paro.

```
desempleo_total <- wb_data("SL.UEM.TOTL.ZS", start_date = 1950, end_date = 2016)
desempleo_total <- tibble(desempleo_total$country,desempleo_total$date,desempleo_total$SL.UEM.TOTL.ZS)
desempleo_total <- desempleo_total %>%
  rename(
    países = "desempleo_total$country",
    fecha = "desempleo_total$date",
    desempleo_t = "desempleo_total$SL.UEM.TOTL.ZS"
  )
desempleo_total <- drop_na(desempleo_total)
desempleo_total_prom <- aggregate(desempleo_total$desempleo_t, list(desempleo_total$países), FUN=mean)
desempleo_total_prom <- desempleo_total_prom %>%
  rename(
    países = Group.1, DESP=x
  )
```

Se incluirá la Tasa promedio de inflación para explicar su efecto parcial sobre el desempeño deportivo en los juegos olímpicos.

```
inflacion <- wb_data("NY.GDP.DEFL.KD.ZG", start_date = 1950, end_date = 2016)
inflacion <- tibble(inflacion$country,inflacion$date,inflacion$NY.GDP.DEFL.KD.ZG)
inflacion <- inflacion %>%
  rename(
    países = "inflacion$country",
    fecha = "inflacion$date",
    inflacion_t = "inflacion$NY.GDP.DEFL.KD.ZG"
  )
inflacion <- drop_na(inflacion)

inflacion_prom <- aggregate(inflacion$inflacion_t, list(inflacion$países), FUN=mean)
inflacion_prom <- inflacion_prom %>%
  rename(
    países = Group.1, INF=x
  )
```

Finalmente, se hace un proceso de adjunción de todas las variables en un mismo objeto.

```
datos <- left_join(medallas_por_pais, growth_pob_prom, by = "países")
View(datos)
datos1 <- left_join(datos, growth_gdp_prom, by = "países")
View(datos1)
datos2 <- left_join(datos1, desempleo_total_prom, by = "países")
View(datos2)
tidy_data <- left_join(datos2, inflacion_prom, by = "países")
View(tidy_data)

loop <- as.data.frame(tidy_data)
for (i in tidy_data) {
  print(class(i))
}
```

```
## [1] "character"
## [1] "integer"
## [1] "numeric"
## [1] "numeric"
## [1] "numeric"
## [1] "numeric"
```

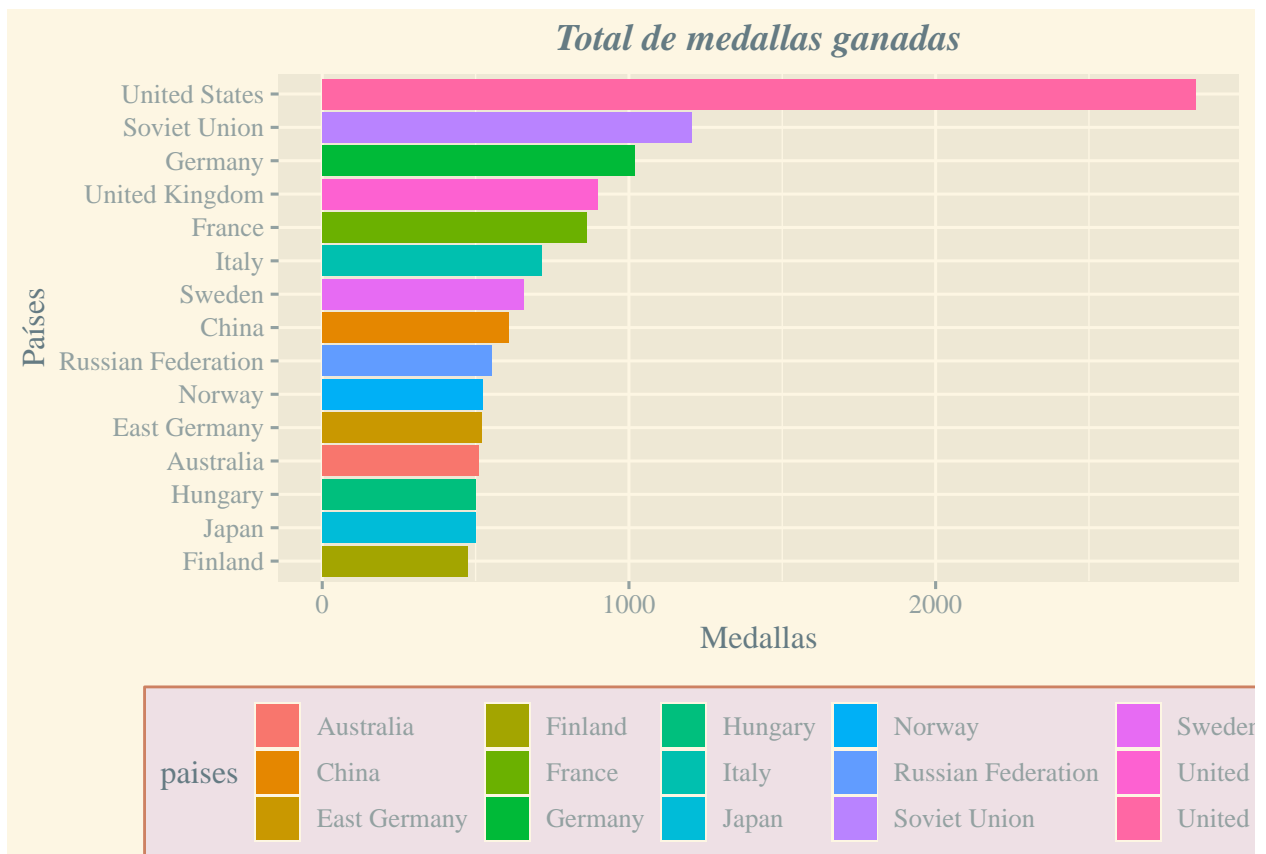
Punto 5

```
resumen_paises <- tidy_data %>%
  summary()
resumen_paises
```

```
##      paises      medallas      POB      GDP
## Length:151      Min.   :  1.0      Min.   : -0.1763      Min.   : -1.976
## Class :character 1st Qu.:  2.0      1st Qu.: 0.5739      1st Qu.: 1.457
## Mode  :character Median : 13.0      Median : 1.5015      Median : 2.222
##              Mean  : 125.9      Mean   : 1.6522      Mean   : 2.300
##              3rd Qu.: 87.5      3rd Qu.: 2.3512      3rd Qu.: 2.931
##              Max.   :2847.0      Max.   : 8.2464      Max.   : 6.848
##              NA's   :17         NA's    :17
##      DESP      INF
## Min.   : 0.4885      Min.   : 1.201
## 1st Qu.: 4.1165      1st Qu.: 4.622
## Median : 7.3373      Median : 7.997
## Mean   : 8.3437      Mean   : 29.657
## 3rd Qu.:11.1727      3rd Qu.: 18.540
## Max.   :33.1550      Max.   :455.599
## NA's   :22          NA's   :17
```

Punto 6

```
tidy_data %>%
  group_by(paises,medallas) %>%
  head(15)%>%
  ggplot(aes(x=reorder(paises,medallas),y=medallas, fill=paises)) +
  geom_col() + coord_flip() + labs(title = "Total de medallas ganadas", x= "Países", y= "Medallas")+
  theme_solarized_2() + scale_colour_solarized('green')+theme(text = element_text(family = "serif"),
plot.title = element_text(face = "bold.italic", hjust = 0.5))+ theme(legend.position = "bottom")+theme
```



Punto 7

```
cor(tidy_data$POB, tidy_data$GDP, use = "complete.obs")
```

```
## [1] -0.3426929
```

```
cor(tidy_data$POB, tidy_data$DESP, use = "complete.obs")
```

```
## [1] -0.2671002
```

```
cor(tidy_data$POB, tidy_data$INF, use = "complete.obs")
```

```
## [1] -0.1541366
```

```
cor(tidy_data$GDP, tidy_data$INF, use = "complete.obs")
```

```
## [1] -0.0718058
```

```
cor(tidy_data$GDP, tidy_data$DESP, use = "complete.obs")
```

```
## [1] 0.0306774
```

```
cor(tidy_data$INF, tidy_data$DESP, use = "complete.obs")
```

```
## [1] 0.1016491
```

```
ggpairs(tidy_data, columns = 3:6, method = c("everything", "pearson"), title = "Correlograma", color = "red",
plot.title = element_text(face = "bold.italic", hjust = 0.5)) + theme(legend.position = "bottom") + theme
```

