

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.

```

gdp_current <- wb_data("NY.GDP.PCAP.CD", start_date = 1950, end_date = 2016)
gdp_current <- tibble(gdp_current$country, gdp_current$date, gdp_current$NY.GDP.PCAP.CD)
gdp_current <- gdp_current %>%
  rename(
    paises = "gdp_current$country",
    fecha = "gdp_current$date",
    usd = "gdp_current$NY.GDP.PCAP.CD"
  )
gdp_current <- drop_na(gdp_current)

gdp_prom <- aggregate(gdp_current$usd, list(gdp_current$paises), FUN=mean)
gdp_prom <- 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")
datos1 <- left_join(datos,gdp_prom,by = "países")
datos2 <- left_join(datos1, desempleo_total_prom,by = "países")
tidy_data<- left_join(datos2, inflacion_prom,by = "países")

tidy_data <- drop_na(tidy_data)
View(tidy_data)

clases <- 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"
```

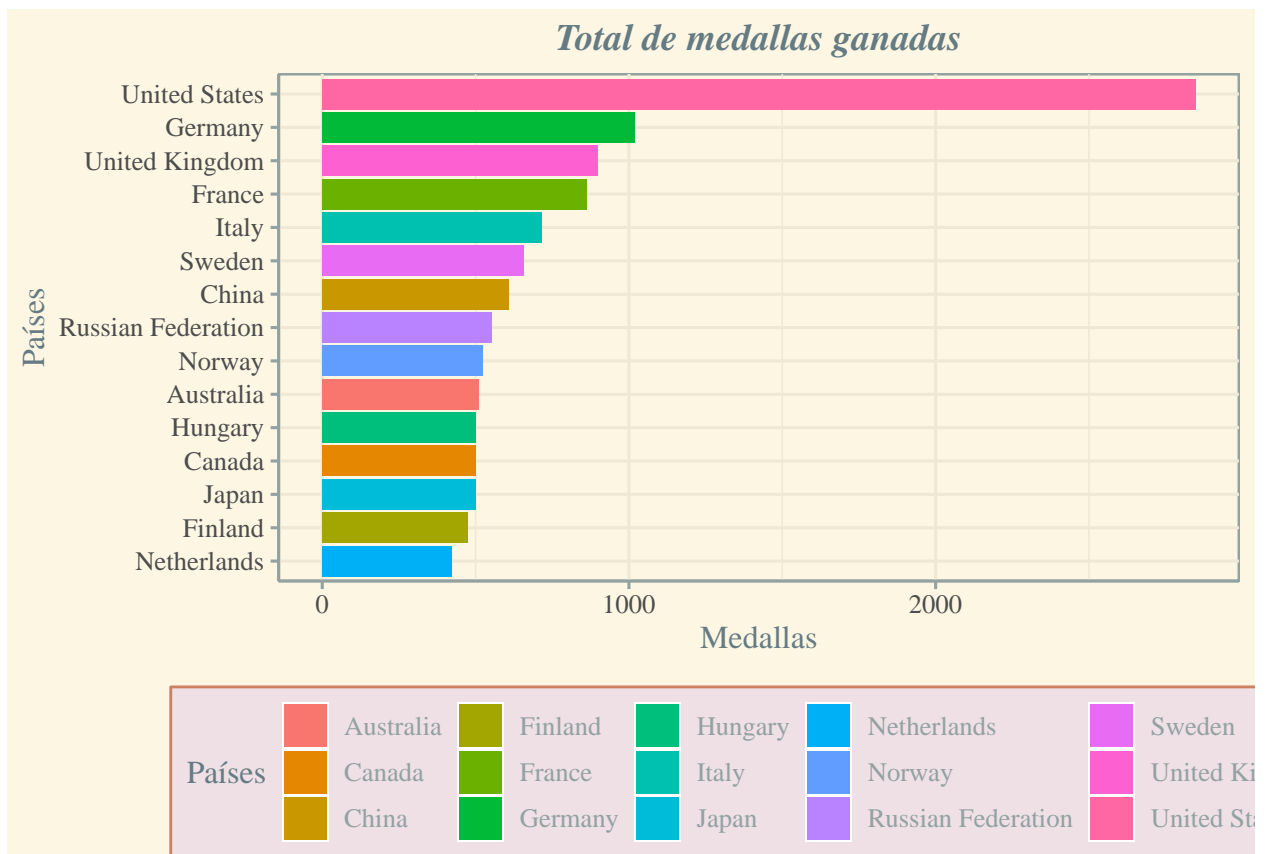
Punto 5

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

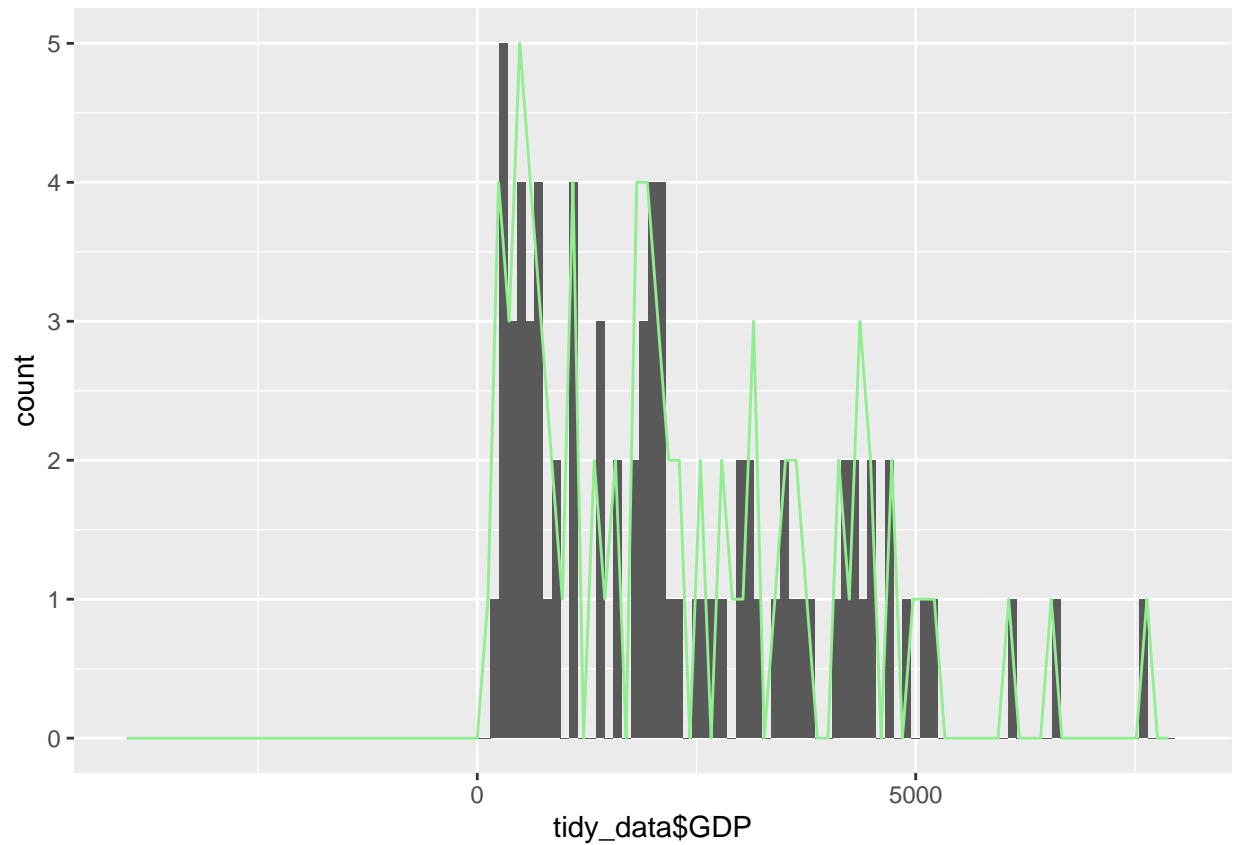
```
##      paises      medallas      POB      GDP
## Length:121      Min.   : 1.0      Min.   :-0.1763      Min.   : 159.5
## Class :character 1st Qu.: 2.0      1st Qu.: 0.5637      1st Qu.: 1561.5
## Mode  :character Median : 15.0      Median : 1.5525      Median : 3681.9
##              Mean  : 135.6      Mean  : 1.6810      Mean  : 7953.6
##              3rd Qu.: 96.0      3rd Qu.: 2.3532      3rd Qu.:12292.5
##              Max.   :2847.0      Max.   : 8.2464      Max.   :40313.6
##      DESP      INF
## Min.   : 0.4885      Min.   : 1.201
## 1st Qu.: 4.0869      1st Qu.: 4.683
## Median : 7.3269      Median : 7.837
## Mean   : 8.3341      Mean   : 28.004
## 3rd Qu.:11.1727      3rd Qu.: 18.194
## Max.   :33.1550      Max.   :455.599
```

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(light = T) + 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
```



```
ggplot(data = tidy_data,
       mapping = aes(x= tidy_data$GDP)) +
  geom_histogram(binwidth = 100) +
  geom_freqpoly(bins=100,color="lightgreen")+
  xlim(-4000, 8000)
```



Punto 7

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

```
## [1] -0.07207423
```

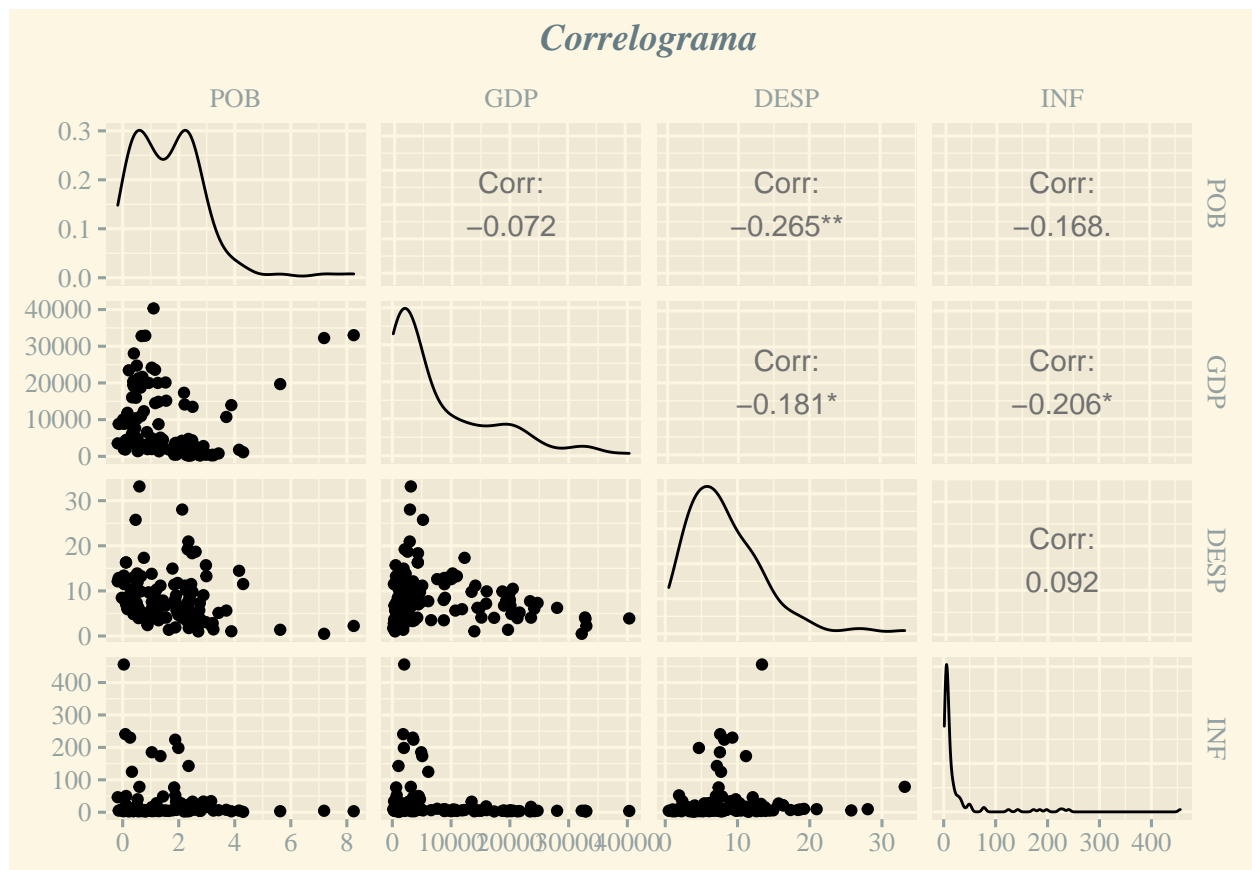
```
cor(tidy_data$POB, tidy_data$DESP, use = "complete.obs")
```

```
## [1] -0.265289
```

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

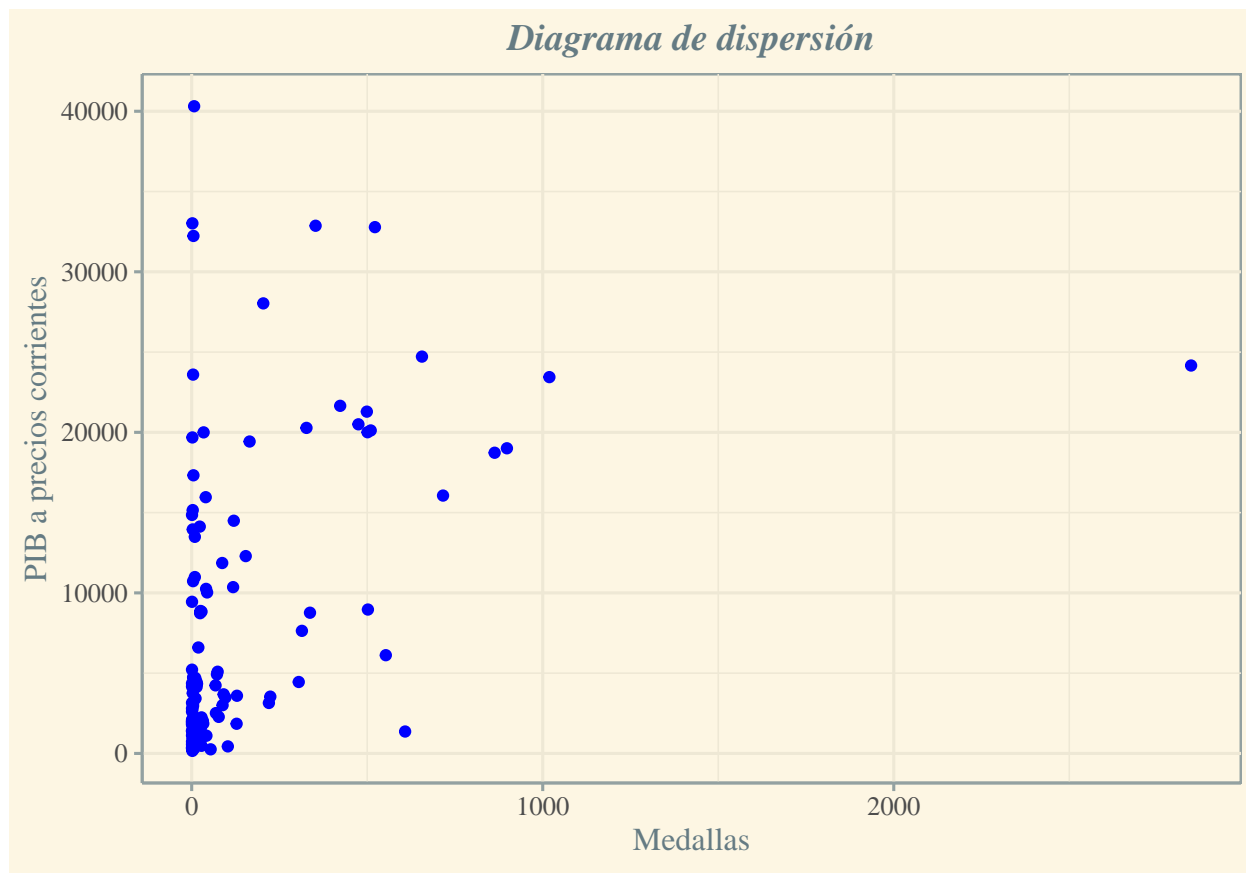
```
## [1] -0.1680504
```

```
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
```

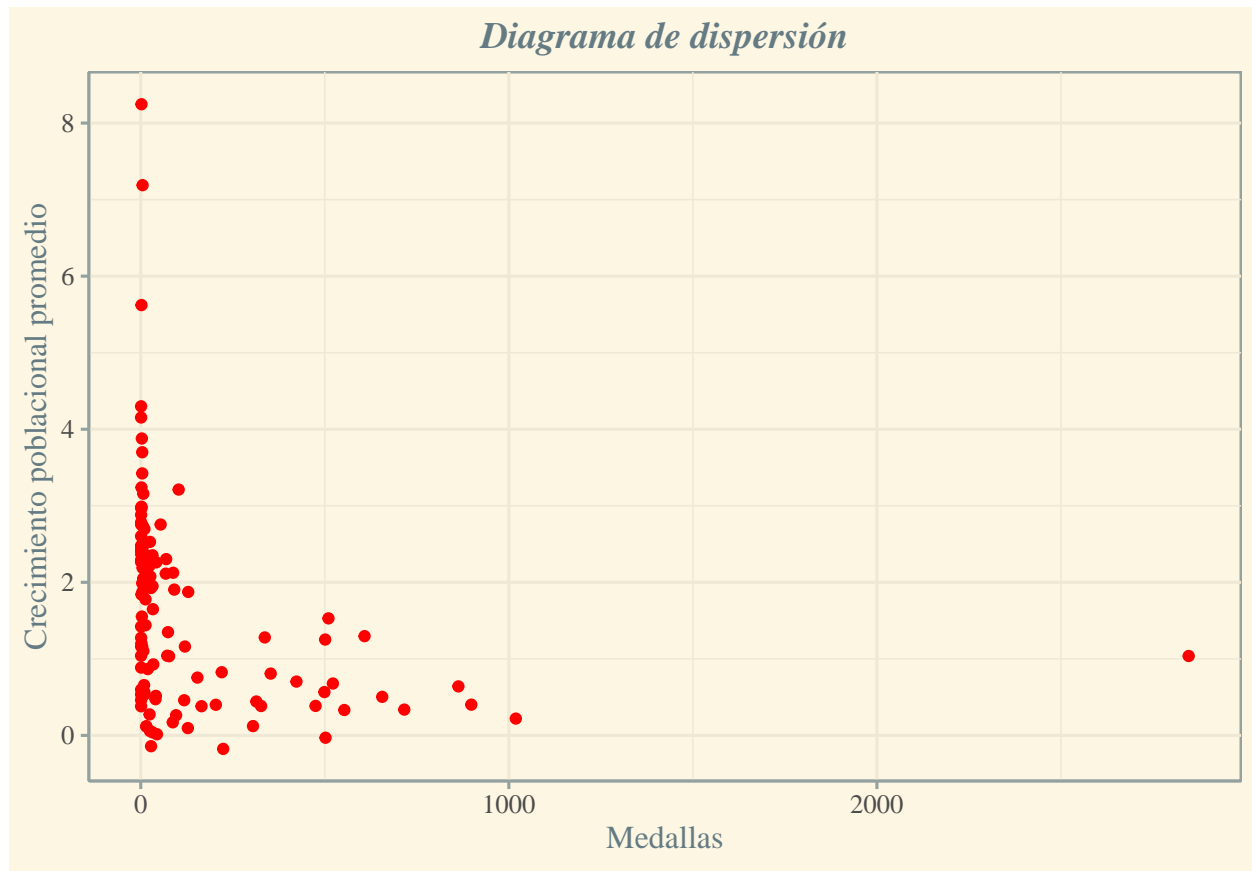


Punto 8

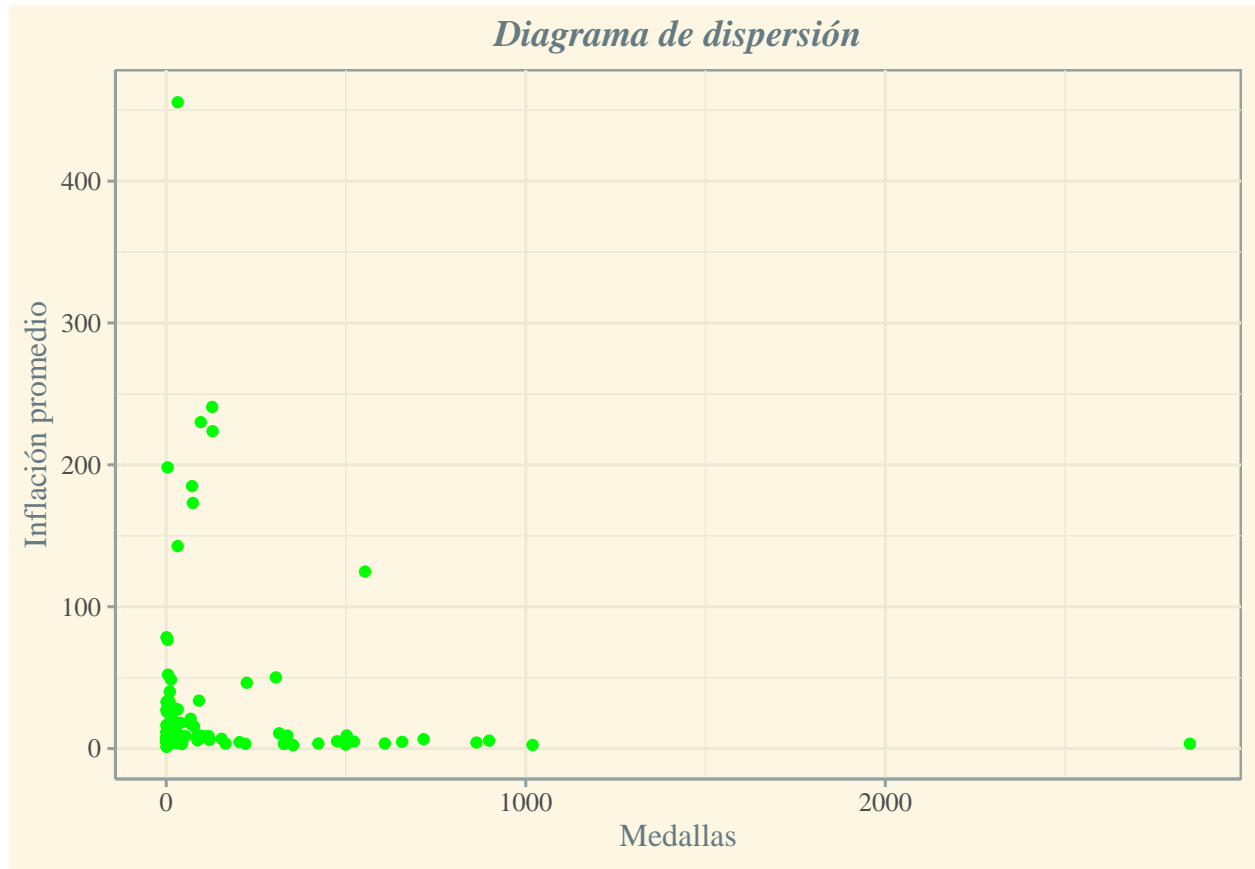
```
ggplot(data = tidy_data, aes(x =medallas,y =GDP)) +
  geom_point(color="blue") +
  labs(title = "Diagrama de dispersión", x= "Medallas", y= "PIB a precios corrientes")+theme_solarized(lig
  plot.title = element_text(face = "bold.italic", hjust = 0.5))+ theme(legend.position = "bottom")+theme
```



```
ggplot(data = tidy_data, aes(x =medallas,y =POB)) +
  geom_point(color="red") +
  labs(title = "Diagrama de dispersión", x= "Medallas", y= "Crecimiento poblacional promedio")+theme_solar
plot.title = element_text(face = "bold.italic", hjust = 0.5))+ theme(legend.position = "bottom")+theme
```

```
ggplot(data = tidy_data, aes(x =medallas,y =INF)) +
  geom_point(color="green") +
  labs(title = "Diagrama de dispersión", x= "Medallas", y= "Inflación promedio")+theme_solarized(light = '
  plot.title = element_text(face = "bold.italic", hjust = 0.5))+ theme(legend.position = "bottom")+theme
```



```
ggplot(data = tidy_data, aes(x =medallas,y =DESP)) +
geom_point(color="black") +
labs(title = "Diagrama de dispersión", x= "Medallas", y= "Desempleo promedio")+theme_solarized(light = '
plot.title = element_text(face = "bold.italic", hjust = 0.5))+ theme(legend.position = "bottom")+theme
```

