

Acceso y uso de World Bank Data mediante WDI

Dereck Amesquita - DAEC Consultoría

Agosto - 2022



Motivación

En este segundo documento de la línea de R y RStudio, se brindará una guía clara sobre el uso de la librería WDI, con la cual podremos acceder a la data oficial del banco mundial. La demanda de datos estadísticos de buena calidad sigue aumentando. Las estadísticas oportunas y confiables son insumos clave para la amplia estrategia de desarrollo. Las mejoras en la calidad y cantidad de datos sobre todos los aspectos del desarrollo son esenciales si queremos lograr el objetivo de un mundo sin pobreza.

¿Por qué usar R en vez del clásico Excel?

Comúnmente podemos descargar esta data de la web del Banco Mundial, pero esto es algo tardado y se complica si queremos usar distintos indicadores para varios países. Veremos que mediante R, este trabajo se vuelve casi automático.

Librerías necesarias :

```
library(tidyverse)
library(WDI)
library(ggplot2)
library(dplyr)
```

Recuerda que si no tienes instaladas estas librerías deberas usar, por ejemplo: `install.packages("WDI")`

La demanda de datos estadísticos de buena calidad sigue aumentando. Las estadísticas oportunas y confiables son insumos clave para la amplia estrategia de desarrollo. Las mejoras en la calidad y cantidad de datos sobre todos los aspectos del desarrollo son esenciales si queremos lograr el objetivo de un mundo sin pobreza.

Sobre la base de datos

En el Banco Mundial, el Grupo de Datos de Desarrollo coordina el trabajo estadístico y de datos y mantiene una serie de bases de datos macro, financieras y sectoriales. Al trabajar en estrecha colaboración con las regiones del Banco y las prácticas globales, el grupo se guía por estándares profesionales en la recopilación, compilación y difusión de datos para garantizar que todos los usuarios de datos puedan tener confianza en la calidad e integridad de los datos producidos. Conoce más en “<https://data.worldbank.org/about>”.

Aclaraciones

R es un software de código abierto, al mismo tiempo este RMarkdown puede ser usado de forma libre.

Resumen

Se iniciara con la descarga de series mediante WDI, posteriormente se enseñara como guardar la data en CSV por si existiera la necesidad. Los primeros gráficos seran expuestos usando el indicador de la inflación. El uso de WDIsearch sera clave para entender como encontrar los códigos de los indicadores que necesitamos. Finalmente buscaremos una serie sobre la esperanza de vida, en la cual haremos un conjunto de tratamiento de datos y de gráficos.

Explorando el potencial de WDI

Asumiremos que ya conocemos los indicadores que necesitamos, en este caso usaremos los siguientes:

- NY.GDP.PCAP.KD.ZG - GDP per capita growth (annual %)
- FP.CPI.TOTL.ZG - Inflation, consumer prices (annual %)

Descarga la data

Creando la variable “datawdi”, obtendremos mediante la función WDI los datos que especifiquemos. En primer lugar, country, en donde podemos crear un vector con las abreviaciones de cada país que necesitemos. Para encontrar la lista de abreviaciones podemos ayudarnos de “https://wits.worldbank.org/wits/wits/witshelp/content/codes/country_codes.htm”. Si quisiéramos la data de todos los países disponibles podríamos usar “all”, en un ejemplo posterior lo veremos. Ahora pasamos a mencionar los indicadores, igual que el anterior, podemos usar un vector, en este caso colocaremos los mencionados anteriormente, si quieres la lista de indicadores tienes 2 opciones, podrías entrar a la web <https://data.worldbank.org/indicator> y ver que cada indicador tiene su código en la URL, la segunda opción y la más profesional es usar WDIsearch, lo cual veremos más adelante.

Tanto “start” como “end” serán usados para establecer las fechas, en el caso de usar una fecha con la que no se cuenten con datos, WDI nos devolverá solo hasta donde se hayan sistematizado dichos datos, es decir, si podemos 2025, nuestra descarga de datos llegará hasta 2021 y no se crearan valores ausentes. Finalmente, language = “es” hará que nos brinde los títulos de los indicadores en español, si no especificamos esto, será en inglés.

```
datawdi = WDI(  
  country = c("PER", "CHL", "COL", "BRA"),  
  indicator = c("NY.GDP.PCAP.KD.ZG", "FP.CPI.TOTL.ZG"),  
  start = 1980,  
  end = 2025,
```

```

language = "es" #Labels en Español
)

str(datawdi)

## 'data.frame': 168 obs. of 5 variables:
## $ iso2c : chr "BR" "BR" "BR" "BR" ...
## $ country : chr "Brasil" "Brasil" "Brasil" "Brasil" ...
## $ year : int 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 ...
## $ NY.GDP.PCAP.KD.ZG: num 6.65 -6.48 -1.5 -5.14 3.06 ...
## $ FP.CPI.TOTL.ZG : num NA 102 101 135 192 ...
## $ attr(*, "label")= chr "Crecimiento del PIB per cápita (% anual)"
## $ attr(*, "label")= chr "Inflación, precios al consumidor (% anual)"

```

```
head(datawdi)
```

```

##   iso2c country year NY.GDP.PCAP.KD.ZG FP.CPI.TOTL.ZG
## 1    BR  Brasil 1980         6.652537             NA
## 2    BR  Brasil 1981        -6.478748         101.7251
## 3    BR  Brasil 1982        -1.503863         100.5434
## 4    BR  Brasil 1983        -5.142491         135.0277
## 5    BR  Brasil 1984         3.063433         192.1217
## 6    BR  Brasil 1985         5.545479         225.9897

```

```
tail(datawdi)
```

```

##   iso2c country year NY.GDP.PCAP.KD.ZG FP.CPI.TOTL.ZG
## 163    PE   Perú 2016         2.4229049         3.557177
## 164    PE   Perú 2017         0.8291263         2.994900
## 165    PE   Perú 2018         2.1981396         1.509154
## 166    PE   Perú 2019         0.6015437         2.252122
## 167    PE   Perú 2020        -12.1987620         2.002412
## 168    PE   Perú 2021         12.0326134         4.271664

```

Renombrar los colnames o los títulos de columnas

Para realizar esta tarea, con fines de práctica, he cambiado los títulos con 3 formas distintas, puedes realizar la que mejor te parezca.

```

names(datawdi)[names(datawdi)=='year'] <- 'Año'
datawdi = datawdi %>% rename (País = country)
datawdi = rename (datawdi, PBIperCapita = NY.GDP.PCAP.KD.ZG)
names(datawdi)[names(datawdi)=='FP.CPI.TOTL.ZG'] <- 'Inflación'

```

Guardar la data en formato CSV

Bastante sencillo si lo que necesitas es llevarte la data a otro lugar, bastará con especificar el DataFrame, en este caso es “datawdi” y posteriormente se le asigna una ubicación.

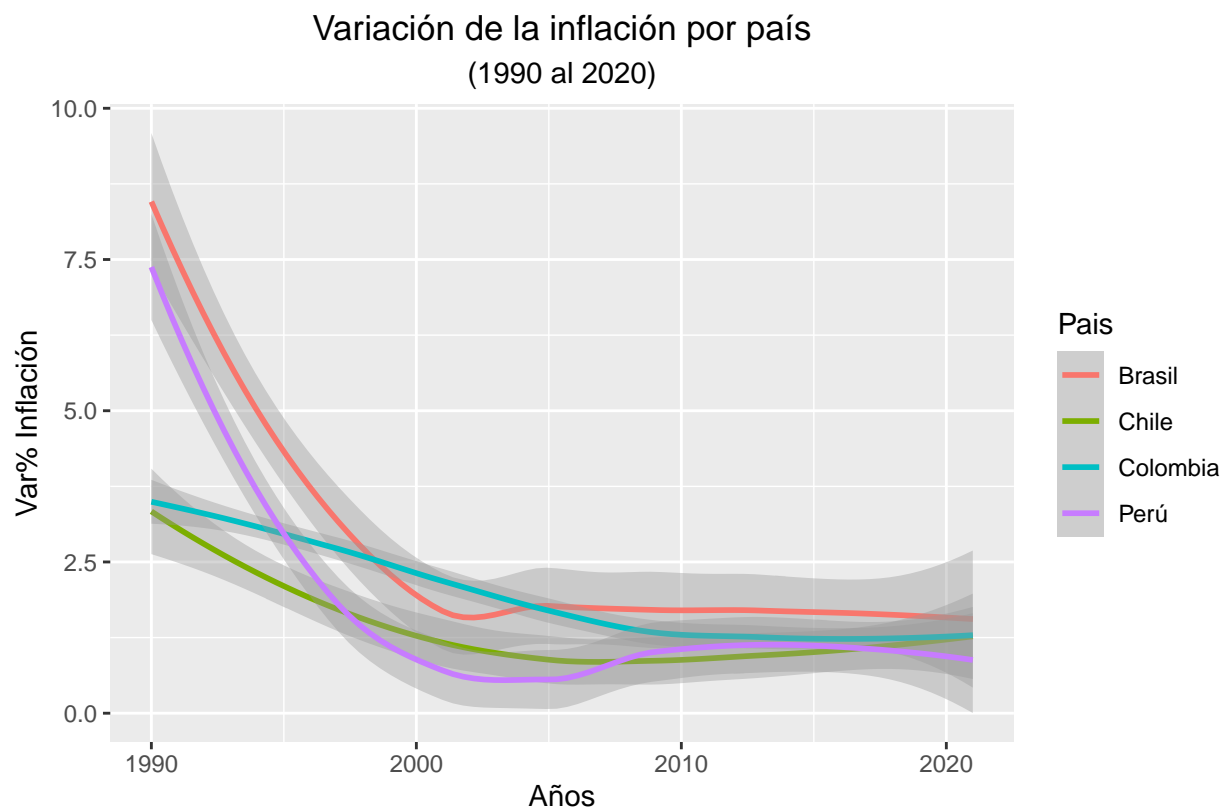
```
write.csv(dataawdi, file="dataawdi.csv")
```

Primeros gráficos

Ahora crearemos unas capas de que le daran cierto estilo a nuestros gráficos con ggplot. Esto lo hago con el fin de no agregar los codigos extensos de los estilos a cada gráfico.

```
ubi = theme(plot.title = element_text(hjust = 0.5), plot.subtitle = element_text(hjust = 0.5), plot.caption = element_text(hjust = 0.5))
tit = labs(caption = "Elaborado por Dereck Amesquita en RStudio")
```

```
grafico1 = ggplot(dataawdi, aes(x=Año, y=log(Inflación), color=País)) + geom_smooth() +
  scale_x_continuous(limits = c(1990, 2021)) +
  labs(x="Años",
       y="Var% Inflación",
       colour= "País",
       title = "Variación de la inflación por país",
       subtitle = "(1990 al 2020)")
grafico1 + ubi + tit # Llamamos a los estilos definidos anteriormente
```



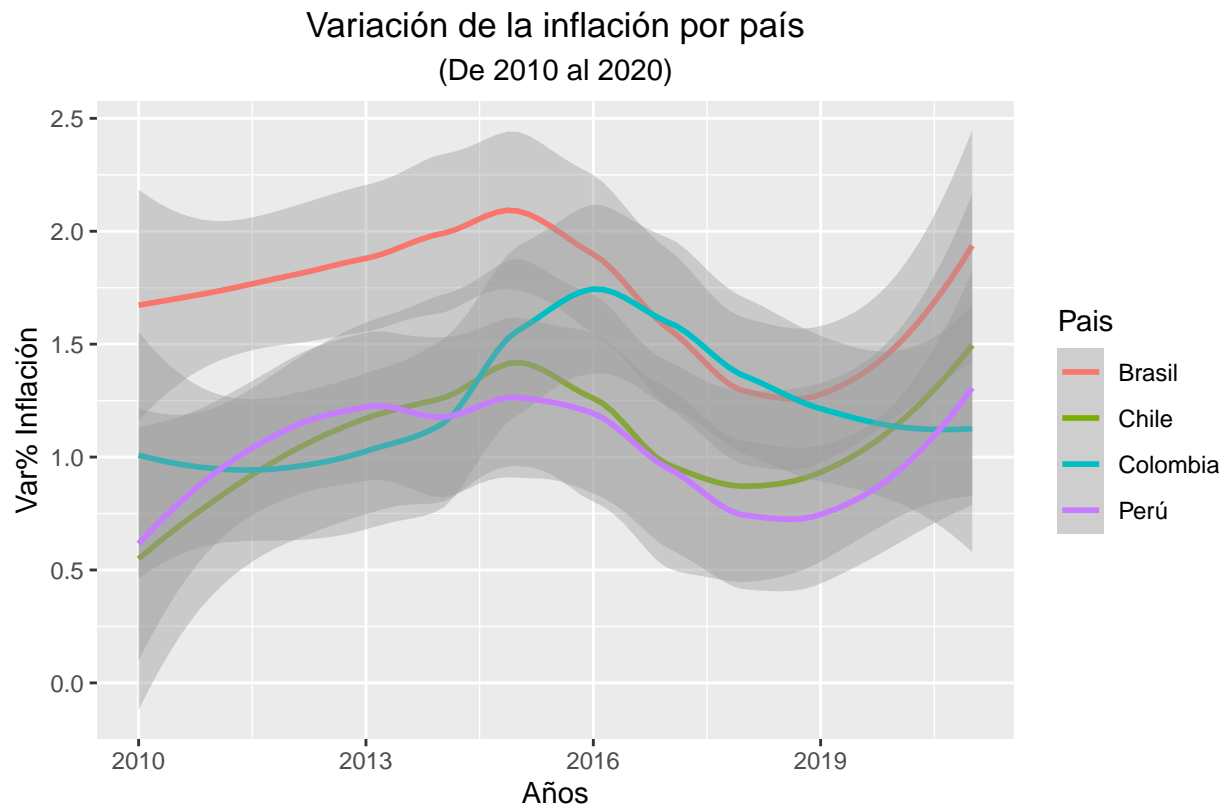
Elaborado por Dereck Amesquita en RStudio

En el caso que quisieramos recortar el número de años podemos hacerlo con “scale_x_continuous”.

```
grafico2 = ggplot(dataawdi, aes(Año, log(Inflación), color=País)) +
  geom_smooth() + scale_x_continuous(limits = c(2010, 2021)) +
  labs(x="Años",
```

```
y="Var% Inflación",
colour= "Pais",
title = "Variación de la inflación por país",
subtitle = "(De 2010 al 2020)"
```

```
grafico2 + ubi + tit # Llamamos a los estilos definidos anteriormente
```



Elaborado por Dereck Amesquita en RStudio

Busqueda de series mediante WDIsearch

Este es el segundo punto más importante, puesto que el Banco Mundial tiene una inmensa cantidad de indicadores y muchas veces podemos perdernos en ello. Usaremos la función “WDIsearch”, mediante string especificaremos alguna palabra clave para encontrar nuestro indicador. Por ejemplo, pondremos “GDP”, es importante resaltar que no hay un buscador en español, y las palabras que usemos tienen que estar en inglés.

```
firstfind = WDIsearch(string='GDP', field='indicator', cache=NULL)
dim(firstfind)
```

```
## [1] 166 2
```

```
head(firstfind)
```

```
## indicator
```

```
## [1,] "5.51.01.10.gdp"
## [2,] "6.0.GDP_current"
## [3,] "6.0.GDP_growth"
## [4,] "6.0.GDP_usd"
## [5,] "6.0.GDPpc_constant"
## [6,] "BN.CAB.XOKA.GDP.ZS"
##      name
## [1,] "Per capita GDP growth"
## [2,] "GDP (current $)"
## [3,] "GDP growth (annual %)"
## [4,] "GDP (constant 2005 $)"
## [5,] "GDP per capita, PPP (constant 2011 international $) "
## [6,] "Current account balance (% of GDP)"
```

La función `dim` nos indica que se tienen 166 observaciones y 2 columnas. Con `head` podemos ver los 6 primeros resultados. Ahora podemos conocer que indicador se ajusta mas a lo que buscamos.

Aplicaciones con la esperanza de vida

Buscaremos “life expectancy at birth”.

```
WDIsearch(string='life expectancy at birth')
```

```
##      indicator      name
## [1,] "SP.DYN.LE00.FE.IN" "Life expectancy at birth, female (years)"
## [2,] "SP.DYN.LE00.IN"    "Life expectancy at birth, total (years)"
## [3,] "SP.DYN.LE00.MA.IN" "Life expectancy at birth, male (years)"
## [4,] "SP.DYN.LIFE.MF"    "Life Expectancy at Birth(years)"
```

Nos quedaremos con el segundo resultado. **Curiosidad: Se podria hacer un estudio diferenciando la esperanza de vida al nacer de hombres y mujeres.** ### Descarga de data En este caso, vamos a descargar la data para todos los paises.

```
data = WDI(
  country = "all",
  indicator = "SP.DYN.LE00.IN",
  start = 2000,
  end = 2025,
  language = "es" #Labels en Español
)
dim(data)
```

```
## [1] 5852    4
```

Tratamiento de datos

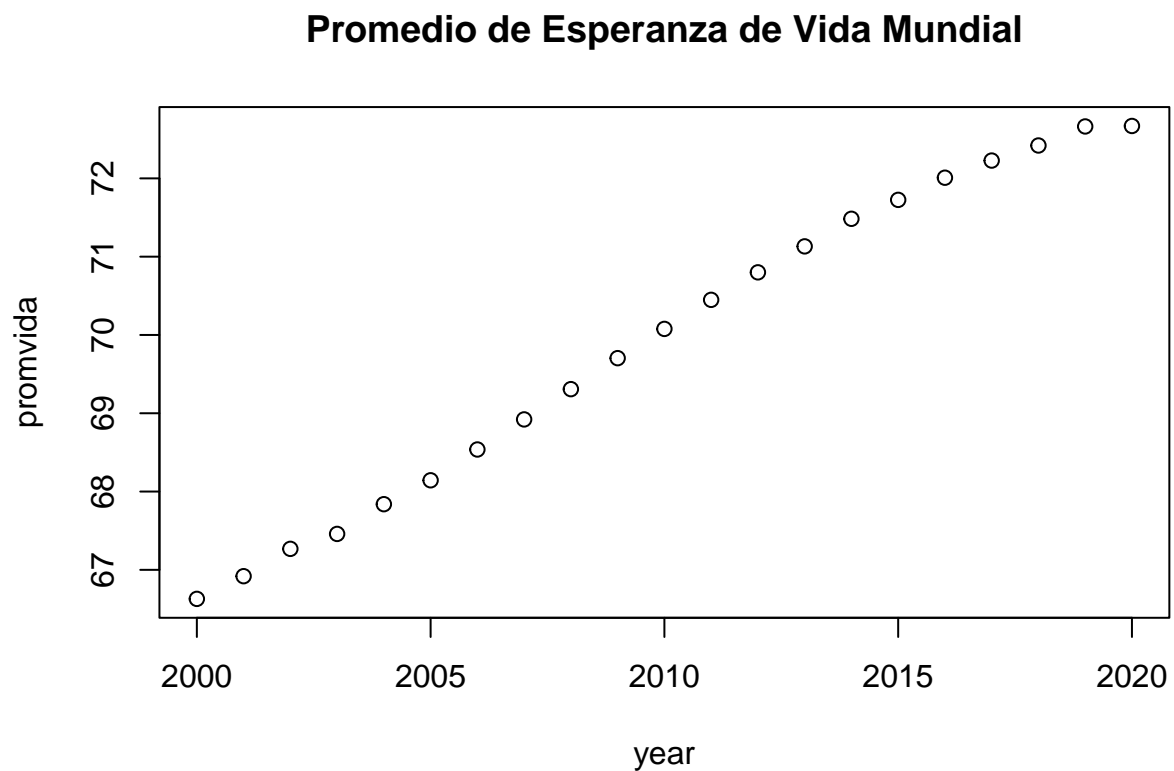
Primero, borraremos todos los datos ausentes. Esto no necesariamente es lo óptimo, pero el objetivo de este documento es el uso de WDI, así que no es del todo necesario centrarnos en el tratamiento de datos.

Crearemos un nuevo DataFrame llamado “esperanza”, el cual sera una agrupación por año de los datos descargados, donde se promediaría la esperanza de vida al nacer. Es decir tendremos una tabla donde esten los años y el promedio de la esperanza de vida de todos los paises.

```
data = na.omit(data)
esperanza = data %>%
  group_by(year) %>%
  summarise(promvida = mean(SP.DYN.LE00.IN))
esperanza
```

```
## # A tibble: 21 x 2
##   year promvida
##   <int>   <dbl>
## 1  2000    66.6
## 2  2001    66.9
## 3  2002    67.3
## 4  2003    67.5
## 5  2004    67.8
## 6  2005    68.1
## 7  2006    68.5
## 8  2007    68.9
## 9  2008    69.3
## 10 2009    69.7
## # ... with 11 more rows
```

```
plot(esperanza, main = "Promedio de Esperanza de Vida Mundial")
```



Países con esperanza de vida mayor a 80 años

Filtraremos dentro de data, pero primero vamos a agrupar a todos por países y ya no años. Entonces

calcularemos la media de la esperanza de vida para cada país a lo largo de los años. Posteriormente filtraremos los que sean mayores e igual a 80, por último usamos arrange para ordenarlos.

```
esp_coun = data %>%
  group_by(country) %>%
  summarise(promvida = mean(SP.DYN.LE00.IN)) %>%
  filter(promvida >=80) %>%
  arrange(desc(promvida))
esp_coun
```

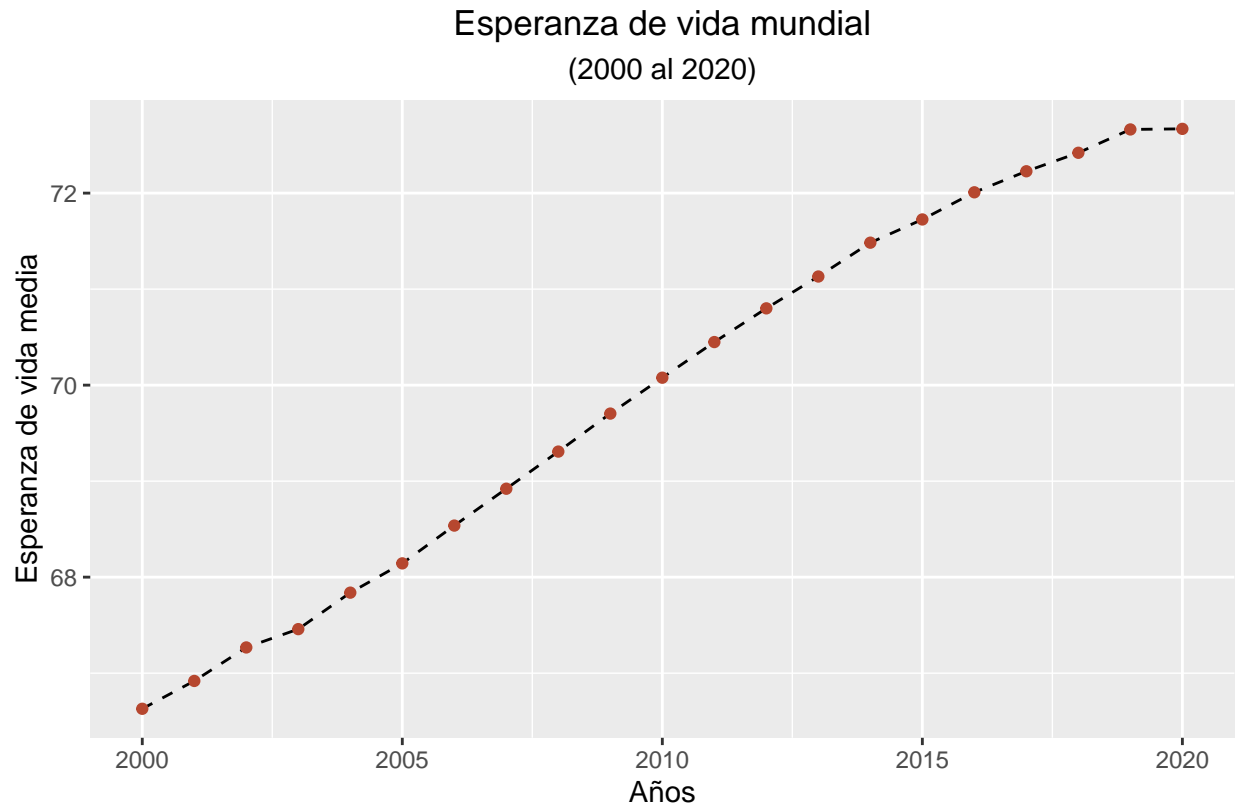
```
## # A tibble: 28 x 2
##   country                promvida
##   <chr>                  <dbl>
## 1 Hong Kong, Región Administrativa Especial    83.1
## 2 Japón                                     82.9
## 3 Región Administrativa Especial de Macao, China 82.6
## 4 Islas Caimán                               82.2
## 5 Suiza                                       82.1
## 6 Islandia                                   81.8
## 7 Italia                                    81.8
## 8 España                                    81.6
## 9 Liechtenstein                             81.6
## 10 Australia                                81.5
## # ... with 18 more rows
```

Gráficos con ggplot

```
esperanza
```

```
## # A tibble: 21 x 2
##   year promvida
##   <int>   <dbl>
## 1 2000    66.6
## 2 2001    66.9
## 3 2002    67.3
## 4 2003    67.5
## 5 2004    67.8
## 6 2005    68.1
## 7 2006    68.5
## 8 2007    68.9
## 9 2008    69.3
## 10 2009    69.7
## # ... with 11 more rows
```

```
grafico3 <- ggplot(data = esperanza, aes(x=year, y=promvida)) +geom_line(linetype = "dashed")+
  geom_point(color = "#B6472F") +
  labs (x="Años" , y ="Esperanza de vida media",
        title = "Esperanza de vida mundial",
        subtitle = "(2000 al 2020)")
grafico3 + ubi + tit
```

Esperanza de vida en Sudamérica

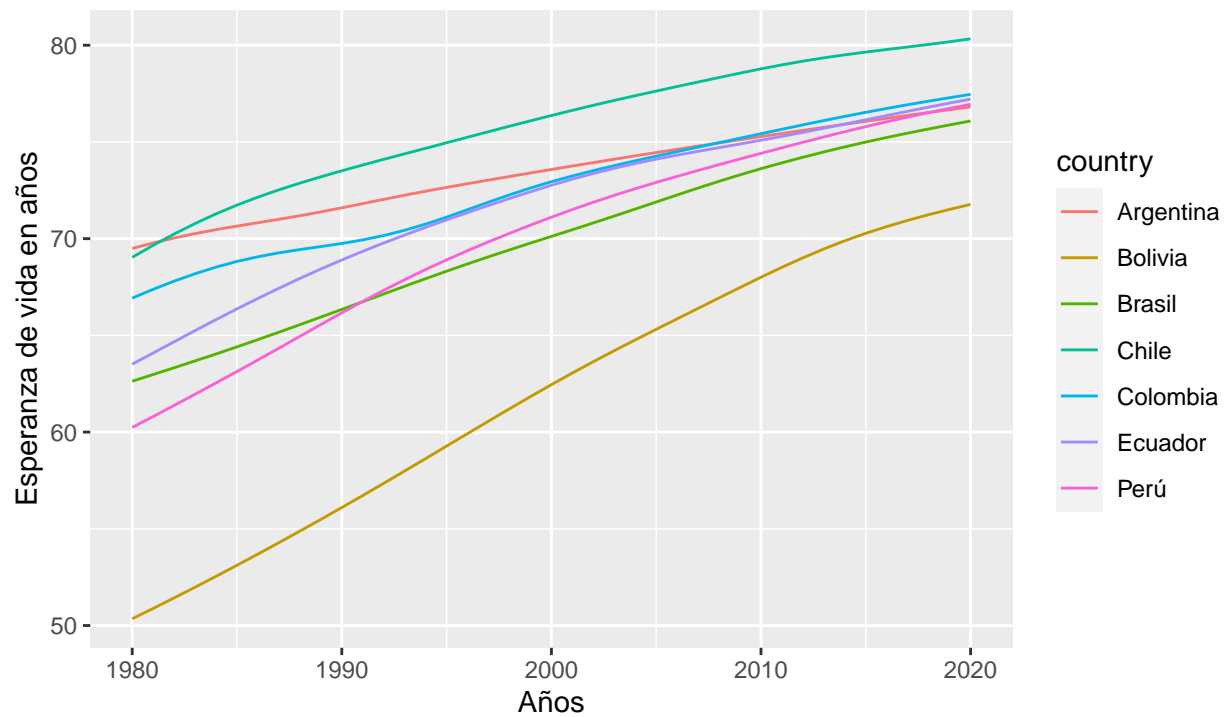
Especificamos los países.

```
data = WDI(
  country = c("PER", "CHL", "COL", "BRA", "ARG", "BOL", "ECU", "COL"),
  indicator = c("SP.DYN.LE00.IN"),
  start = 1980,
  end = 2025,
  language = "es" #Labels en Español
)

data = na.omit(data)

grafico4 <- ggplot(data, aes(x=year, y=SP.DYN.LE00.IN, color=country)) + geom_line() +
  labs(x="Años", y = "Esperanza de vida en años",
       title = "Esperanza de vida por países",
       subtitle = "(De 1980 a 2020)")
grafico4 + ubi + tit
```

Esperanza de vida por países (De 1980 a 2020)

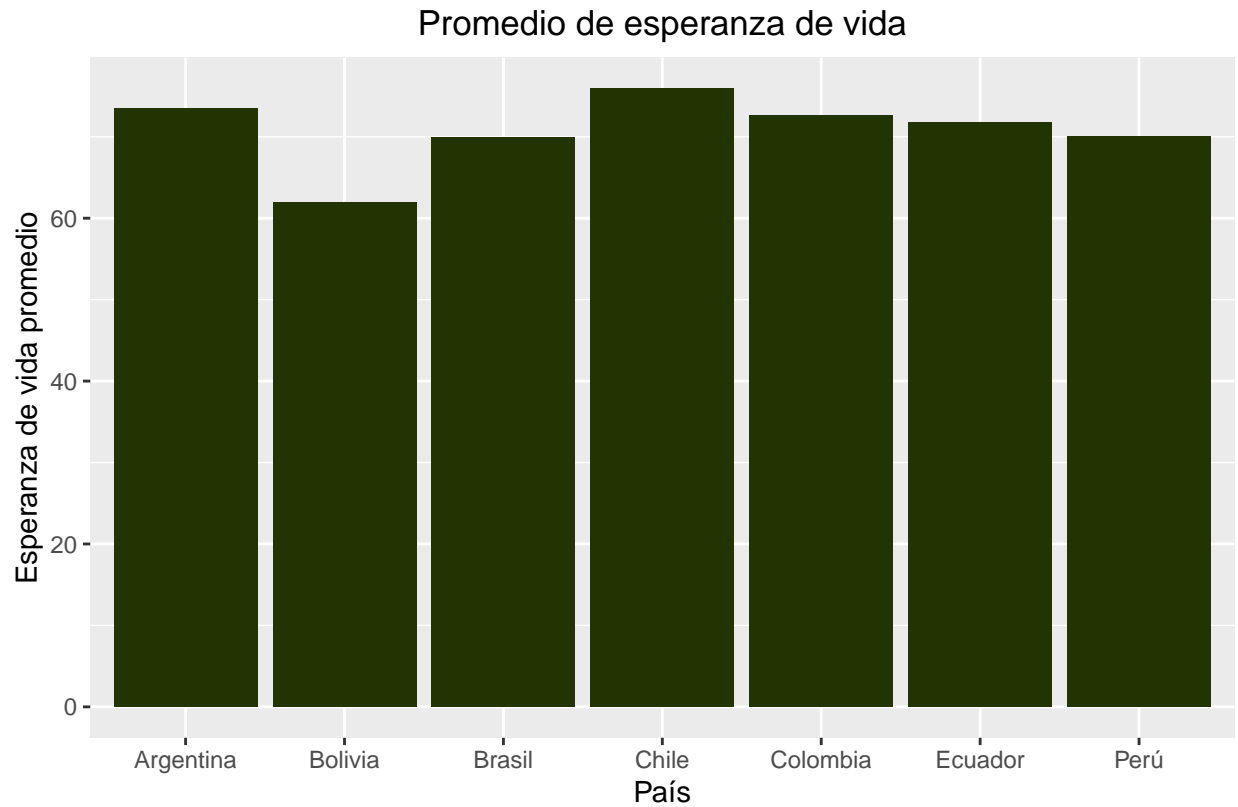


Elaborado por Dereck Amesquita en RStudio

```
esperanza = data %>%
  group_by(country) %>%
  summarise(promvida = mean(SP.DYN.LE00.IN))
esperanza
```

```
## # A tibble: 7 x 2
##   country promvida
##   <chr>      <dbl>
## 1 Argentina    73.4
## 2 Bolivia      61.9
## 3 Brasil       69.9
## 4 Chile        75.9
## 5 Colombia     72.6
## 6 Ecuador      71.8
## 7 Perú         70.1
```

```
grafico5 = ggplot(esperanza, aes(x=country, y=promvida)) +
  geom_col(fill='#213402') +
  labs(x='País', y="Esperanza de vida promedio", title = 'Promedio de esperanza de vida')
grafico5+ubi+tit
```



Agradecimiento

Este documento fue realizado en parte por la inspiración al leer un Rmarkdown publicado por **Ramfis Felix Tatis (Economista)**. Link de la versión PDF(t.ly/z_Cu), también aprovecho para agradecerle por la ayuda brindada para usar WDIsearch y encontrar el código de los indicadores.

Enlaces importantes

- <https://data.worldbank.org/indicator>
- <https://datahelpdesk.worldbank.org/knowledgebase/articles/201175-how-does-the-world-bank-code-its-indicators>
- <https://cran.r-project.org/web/packages/WDI/WDI.pdf>
- https://wits.worldbank.org/wits/wits/witshelp/content/codes/country_codes.htm