Alfredo Espioza (20170552)

Expectativa de vida al nacer en años. Se usará los datos del 2017

Traemos los datos de Github

linkvida="https://github.com/Alfredo-Espinoza/Estadistica-2-Parte-individual-del-trabajo-grupal/raw/master/API\_SP.DYN.LE00.IN\_DS2\_en\_excel\_v2\_199538.xlsx"  
#Descargando en carpeta  
download.file(linkvida, "vida.xlsx")

Unimos todas las hojas:

library(rio)  
basevida <- import\_list('vida.xlsx',rbind = TRUE)

## New names:  
## \* `` -> ...3  
## \* `` -> ...4  
## \* `` -> ...5  
## \* `` -> ...6  
## \* `` -> ...7  
## \* … and 56 more problems

Borrando columnas innecesarias

basevida[2:50]=NULL

basevida[2:12]=NULL

basevida[3:13]=NULL

Revisamos la base de datos

str(basevida)

## 'data.frame': 531 obs. of 2 variables:  
## $ Data Source: chr "Last Updated Date" NA "Country Name" "Aruba" ...  
## $ ...62 : chr NA NA "2017" "76.010000000000005" ...  
## - attr(\*, ".internal.selfref")=<externalptr>

Borramos los NAs

basevida=basevida[complete.cases(basevida),]

Cambiando nombres de variable

names(basevida)[1]="Pais"  
names(basevida)[2]="Expectativa de vida en años al nacer (2017)"

Eliminamos la primera fila, porque es producto de la suciedad de losd atos.

basevida=basevida[-c(1),]

Eliminando posibles espacios vacios

basevida$Pais=trimws(basevida$Pais,whitespace = "[\\h\\v]")  
basevida$`Expectativa de vida en años al nacer (2017)`=trimws(basevida$`Expectativa de vida en años al nacer (2017)`,whitespace = "[\\h\\v]")

Revisamos la base de datos

str(basevida)

## 'data.frame': 245 obs. of 2 variables:  
## $ Pais : chr "Aruba" "Afghanistan" "Angola" "Albania" ...  
## $ Expectativa de vida en años al nacer (2017): chr "76.010000000000005" "64.13" "60.378999999999998" "78.332999999999998" ...  
## - attr(\*, ".internal.selfref")=<externalptr>

Pasando a numérico la variable a estudiar

basevida$"Expectativa de vida en años al nacer (2017)"=as.numeric(basevida$"Expectativa de vida en años al nacer (2017)")

str(basevida)

## 'data.frame': 245 obs. of 2 variables:  
## $ Pais : chr "Aruba" "Afghanistan" "Angola" "Albania" ...  
## $ Expectativa de vida en años al nacer (2017): num 76 64.1 60.4 78.3 71.6 ...  
## - attr(\*, ".internal.selfref")=<externalptr>

Hay observaciones que corresponden a la suciedad de la data. Varias de ellas serán eliminadas una vez se realiza el merge de acuerdo a “Pais” con las variables del grupo de trabajo. El resto tendrán que ser eliminadas manualmente.