Alfredo Espinoza (20170552) Porcentaje de uso de internet por país. Se usarán los datos del 2017

Trayendo desde Github

linkinter="https://github.com/Alfredo-Espinoza/Estadistica-2-Parte-individual-del-trabajo-grupal/raw/master/API\_IT.NET.USER.ZS\_DS2\_en\_excel\_v2\_199844.xlsx"  
#Descargando en carpeta  
download.file(linkinter, "intercel.xlsx")

Unimos todas las hojas:

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

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

Revisamos la base

str(internet)

## 'data.frame': 531 obs. of 73 variables:  
## $ Data Source : chr "Last Updated Date" NA "Country Name" "Aruba" ...  
## $ World Development Indicators: chr "43735" NA "Country Code" "ABW" ...  
## $ ...3 : chr NA NA "Indicator Name" "Individuals using the Internet (% of population)" ...  
## $ ...4 : chr NA NA "Indicator Code" "IT.NET.USER.ZS" ...  
## $ ...5 : chr NA NA "1960" NA ...  
## $ ...6 : chr NA NA "1961" NA ...  
## $ ...7 : chr NA NA "1962" NA ...  
## $ ...8 : chr NA NA "1963" NA ...  
## $ ...9 : chr NA NA "1964" NA ...  
## $ ...10 : chr NA NA "1965" NA ...  
## $ ...11 : chr NA NA "1966" NA ...  
## $ ...12 : chr NA NA "1967" NA ...  
## $ ...13 : chr NA NA "1968" NA ...  
## $ ...14 : chr NA NA "1969" NA ...  
## $ ...15 : chr NA NA "1970" NA ...  
## $ ...16 : chr NA NA "1971" NA ...  
## $ ...17 : chr NA NA "1972" NA ...  
## $ ...18 : chr NA NA "1973" NA ...  
## $ ...19 : chr NA NA "1974" NA ...  
## $ ...20 : chr NA NA "1975" NA ...  
## $ ...21 : chr NA NA "1976" NA ...  
## $ ...22 : chr NA NA "1977" NA ...  
## $ ...23 : chr NA NA "1978" NA ...  
## $ ...24 : chr NA NA "1979" NA ...  
## $ ...25 : chr NA NA "1980" NA ...  
## $ ...26 : chr NA NA "1981" NA ...  
## $ ...27 : chr NA NA "1982" NA ...  
## $ ...28 : chr NA NA "1983" NA ...  
## $ ...29 : chr NA NA "1984" NA ...  
## $ ...30 : chr NA NA "1985" NA ...  
## $ ...31 : chr NA NA "1986" NA ...  
## $ ...32 : chr NA NA "1987" NA ...  
## $ ...33 : chr NA NA "1988" NA ...  
## $ ...34 : chr NA NA "1989" NA ...  
## $ ...35 : chr NA NA "1990" "0" ...  
## $ ...36 : chr NA NA "1991" NA ...  
## $ ...37 : chr NA NA "1992" NA ...  
## $ ...38 : chr NA NA "1993" NA ...  
## $ ...39 : chr NA NA "1994" NA ...  
## $ ...40 : chr NA NA "1995" NA ...  
## $ ...41 : chr NA NA "1996" "2.7683826630000001" ...  
## $ ...42 : chr NA NA "1997" NA ...  
## $ ...43 : chr NA NA "1998" NA ...  
## $ ...44 : chr NA NA "1999" "4.5061790979999996" ...  
## $ ...45 : chr NA NA "2000" "15.44282295" ...  
## $ ...46 : chr NA NA "2001" "17.100000000000001" ...  
## $ ...47 : chr NA NA "2002" "18.8" ...  
## $ ...48 : chr NA NA "2003" "20.8" ...  
## $ ...49 : chr NA NA "2004" "23" ...  
## $ ...50 : chr NA NA "2005" "25.4" ...  
## $ ...51 : chr NA NA "2006" "28" ...  
## $ ...52 : chr NA NA "2007" "30.9" ...  
## $ ...53 : chr NA NA "2008" "52" ...  
## $ ...54 : chr NA NA "2009" "58" ...  
## $ ...55 : chr NA NA "2010" "62" ...  
## $ ...56 : chr NA NA "2011" "69" ...  
## $ ...57 : chr NA NA "2012" "74" ...  
## $ ...58 : chr NA NA "2013" "78.900000000000006" ...  
## $ ...59 : chr NA NA "2014" "83.78" ...  
## $ ...60 : chr NA NA "2015" "88.661226929999998" ...  
## $ ...61 : chr NA NA "2016" "93.542453870000003" ...  
## $ ...62 : chr NA NA "2017" "97.17" ...  
## $ ...63 : chr NA NA "2018" NA ...  
## $ \_file : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Country Code : chr NA NA NA NA ...  
## $ Region : chr NA NA NA NA ...  
## $ IncomeGroup : chr NA NA NA NA ...  
## $ SpecialNotes : chr NA NA NA NA ...  
## $ TableName : chr NA NA NA NA ...  
## $ INDICATOR\_CODE : chr NA NA NA NA ...  
## $ INDICATOR\_NAME : chr NA NA NA NA ...  
## $ SOURCE\_NOTE : chr NA NA NA NA ...  
## $ SOURCE\_ORGANIZATION : chr NA NA NA NA ...  
## - attr(\*, ".internal.selfref")=<externalptr>

Borrando las columnas de otros años

internet[2:50]=NULL

internet[2:12]=NULL

internet[3:13]=NULL

Eliminando los NA

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

Cambiando nombres de variables.

names(internet)[1]="Pais"

names(internet)[2]="Porcentaje de la población que usa internet (2017)"

Eliminando posibles espacios vacios

internet$Pais=trimws(internet$Pais,whitespace = "[\\h\\v]")  
internet$`Porcentaje de la población que usa internet (2017)`=trimws(internet$`Porcentaje de la población que usa internet (2017)`,whitespace = "[\\h\\v]")

Eliminando primera fila.

internet=internet[-1,]

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.

Pasando la variable a numérico

internet$`Porcentaje de la población que usa internet (2017)`=as.numeric(internet$`Porcentaje de la población que usa internet (2017)`)

Revisando

str(internet)

## 'data.frame': 253 obs. of 2 variables:  
## $ Pais : chr "Aruba" "Afghanistan" "Angola" "Albania" ...  
## $ Porcentaje de la población que usa internet (2017): num 97.2 13.5 14.3 71.8 91.6 ...  
## - attr(\*, ".internal.selfref")=<externalptr>