

# Tarea TidyR

Análisis Exploratorio de Datos, Máster en Ciencia de Datos- UV

Daniel Lillo Plaza

2023-07-09

## Índice

1	Introducción.	1
2	Paquetes y librerías necesarios.	1
3	Tarea.	2

---

## 1 Introducción.

El objetivo de esta tarea es practicar la obtención de formatos de datos *tidy* a partir de datos no estructurados.

## 2 Paquetes y librerías necesarios.

Incluimos todas las librerías necesarias para la ejecución del código en la siguiente lista. `packages = c(tidyr,"knitr",...)`

Si la librería no está instalada se instalará y cargará, si no, solo se cargará.

```
# Especificamos las librerías necesarias en esta lista

packages = c("tidyr","knitr", "dplyr")

#use this function to check if each package is on the local machine
#if a package is installed, it will be loaded
#if any are not, the missing package(s) will be installed and loaded
package.check <- lapply(packages, FUN = function(x) {
  if (!require(x, character.only = TRUE)) {
    install.packages(x, dependencies = TRUE)
    library(x, character.only = TRUE)
  }
})

#verify they are loaded
search()
```

```
[1] ".GlobalEnv"      "package:dplyr"    "package:tidyr"    "package:knitr"
[5] "package:stats"    "package:graphics" "package:grDevices" "package:utils"
[9] "package:datasets" "package:methods"  "Autoloads"        "package:base"
```

### 3 Tarea.

En la librería tidyR hay varios data frame de datos: `table1`, `table2`, `table3`, `table4a`, `table4b`, `table5`. `Table1`, presenta una estructura de datos “tidy”, pero el resto no.

Escribe un proyecto basado en el proyecto plantilla que realice las siguientes tareas:

1. Observa la estructura de todos los data frames y determina qué tipo de anomalía presenta para considerarlo un fichero “tidy”.

Cuadro 1: Table1

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

Cuadro 2: Table2

country	year	type	count
Afghanistan	1999	cases	745
Afghanistan	1999	population	19987071
Afghanistan	2000	cases	2666
Afghanistan	2000	population	20595360
Brazil	1999	cases	37737
Brazil	1999	population	172006362
Brazil	2000	cases	80488
Brazil	2000	population	174504898
China	1999	cases	212258
China	1999	population	1272915272
China	2000	cases	213766
China	2000	population	1280428583

Cuadro 3: Table3

country	year	rate
Afghanistan	1999	745/19987071
Afghanistan	2000	2666/20595360
Brazil	1999	37737/172006362
Brazil	2000	80488/174504898
China	1999	212258/1272915272

country	year	rate
China	2000	213766/1280428583

Cuadro 4: Table4a

country	1999	2000
Afghanistan	745	2666
Brazil	37737	80488
China	212258	213766

Cuadro 5: Table4b

country	1999	2000
Afghanistan	19987071	20595360
Brazil	172006362	174504898
China	1272915272	1280428583

Cuadro 6: Table5

country	century	year	rate
Afghanistan	19	99	745/19987071
Afghanistan	20	00	2666/20595360
Brazil	19	99	37737/172006362
Brazil	20	00	80488/174504898
China	19	99	212258/1272915272
China	20	00	213766/1280428583

Observamos que `table2` no es tidy

2. Obtén `table1` a partir de `table2`.

```
table2_new<-pivot_wider(table2, names_from="type", values_from="count")
kable(table2_new)
```

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

3. Obtén `table1` a partir de `table3`.

```
table3_new<-separate(table3, rate, c("cases", "population"), sep="/")
kable(table3_new)
```

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

4. Obtén table1 a partir de table4a y table4b.

```
table4a_new<-pivot_longer(table4a, names_to="year", values_to="cases", cols=c(2,3))
table4b_new<-pivot_longer(table4b, names_to="year", values_to="population", cols=c(2,3))
table4ab_new<-full_join(table4a_new, table4b_new)
kable(table4ab_new)
```

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

5. Obtén table1 a partir de table5.

```
table5_new<-unite(table5, year, century, year, sep="")
table5_new<-separate(table5_new, rate, c("cases", "population"), sep="/")
kable(table5_new)
```

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583