

LAB 29 (MD)

AnaGSanjuanM

2023-02-23

LABORATORIO 29

Uso de herramienta dplyr

Sumar una cierta variable pero agrupado por otra variable (ejemplo PCI con location_name)

Cargar paquetes

```
library(data.table)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##      between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

Cargar base de datos choose.files() seleccionando la carpeta

```
Creamos objeto llamado data2013

Colocamos asignación (alt -) <-

El archivo se leerá con read.csv y se copia y pega la ruta del archivo

```r
data2013 <- read.csv("C:\\Users\\Lenovo\\Documents\\GitHub\\LAB-29\\L29 input\\greenR_2013.csv")
```

Se transforma a tabla de datos

Signo de asignación y debe leerlo as.data.table de data 2013

```
data2013 <- as.data.table(data2013)
```

Para saber qué clase de datos son se coloca class

```
class(data2013)
```

```
[1] "data.table" "data.frame"
```

En la consola se puede visualizar que es un datatable, también llamad dataframe

Suma de variable agrupada por la categoría de la entidad

Crear dataframe resultados. Signo de asignación <- Proviene de data2013 y activamos operador pipe

Se agrupan group\_by por el nombre de la localidad

Otro pipe %>% para el summarise (la suma) de GCI\_2013 que será igual a la suma de dicha variable

na.rm significa que los valores faltantes se omitan, por lo que colocamos TRUE.

También se requiere otra suma de GP\_SINVCR.

El último summarise es la variable GCP\_2013 omitiendo valores faltantes.

Por último, hacer conteo de las celdas

```
resultados2013 <- data2013 %>%
 group_by(location_name) %>%
 summarise(suma_GCI_2013 = sum(GCI_2013, na.rm = T),
 suma_GP_SINVCR = sum(GP_SINVCR, na.rm = T),
 suma_GCP_2013 = sum(GCP_2013, na.rm = T),
 count = n())
```

Para visualizar

```
resultados2013
```

```
A tibble: 32 × 5
location_name suma_GCI_2013 suma_GP_SINVCR suma_GCP_2013 count
<chr> <dbl> <int> <dbl> <int>
1 Aguascalientes 16.1 91 2.98 106
2 Baja California 25.7 75 6.34 106
3 Baja California Sur 2.74 101 1.58 106
4 Campeche -0.217 105 0.280 106
5 Chiapas 1.22 103 1.12 106
6 Chihuahua 17.9 83 5.19 106
7 Coahuila 18.6 79 5.22 106
8 Colima -2.34 101 1.76 106
9 Distrito Federal 6.24 97 7.27 106
10 Durango 8.34 93 3.26 106
... with 22 more rows
```