

13. -Data Visualization_04_19_vacunacion_completo_v_01

June 8, 2023

#

CU04_Optimización de vacunas

Citizenlab Data Science Methodology > II - Data Processing Domain *** > # 13.- Data Visualization

Data Visualization is the process of performing a statistical graphical analysis of the data.

0.1 Notas

- La visualización es parte del análisis exploratorio de datos
- En los notebooks del proceso 12 se incluyen gráficos que constituyen la visualización de los datos del caso de uso, y no se repiten aquí
- En los notebooks del proceso 13 se añaden algunas visualizaciones espaciales no incluidas en el proceso 12

0.2 File

En este notebook se importan dos archivos: el csv con los datos y el json con las geometrías

- Input File: CU_04_08_20_vacunacion_gripe_train_and_test.csv
- No aplica

0.2.1 Encoding

Con la siguiente expresión se evitan problemas con el encoding al ejecutar el notebook. Es posible que deba ser eliminada o adaptada a la máquina en la que se ejecute el código.

```
[1]: Sys.setlocale(category = "LC_ALL", locale = "es_ES.UTF-8")
```

```
Warning message in Sys.setlocale(category = "LC_ALL", locale = "es_ES.UTF-8"):  
"OS reports request to set locale to "es_ES.UTF-8" cannot be honored"
```

```
"
```

0.3 Settings

0.3.1 Libraries to use

```
[2]: library(readr)
library(dplyr)
library(tidyr)
library(ggplot2)
library(summarytools)
library(GGally)
library(sf)
library(nortest)
library(lubridate)
library(leaflet)
```

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

filter, lag

The following objects are masked from ‘package:base’:

intersect, setdiff, setequal, union

Registered S3 method overwritten by 'GGally':

method from
+.gg ggplot2

Linking to GEOS 3.10.2, GDAL 3.4.1, PROJ 8.2.1; sf_use_s2() is TRUE

Attaching package: ‘lubridate’

The following objects are masked from ‘package:base’:

date, intersect, setdiff, union

0.3.2 Paths

```
[3]: iPath <- "Data/Input/"  
     oPath <- "Data/Output/"
```

0.4 Data Load

1. Archivo de datos CSV

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Uncomment the line if using this option

```
[4]: # file_data <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

```
[5]: iFile <- "CU_04_08_20_vacunacion_gripe_train_and_test.csv"  
     file_data <- paste0(iPath, iFile)  
  
     if(file.exists(file_data)){  
       cat("Se leerán datos del archivo: ", file_data)  
     } else{  
       warning("Cuidado: el archivo no existe.")  
     }  
}
```

Se leerán datos del archivo:

Data/Input/CU_04_08_20_vacunacion_gripe_train_and_test.csv

Data file to dataframe Usar la función adecuada según el formato de entrada (xlsx, csv, json, ...)

```
[6]: data <- read_csv(file_data)
```

Rows: 21736 Columns: 49

Column specification

Delimiter: ","

chr (3): GEOCODIGO, DESBDT, nombre_zona

dbl (45): ano, semana, n_vacunas, n_citas, tmed, prec, velmedia,
presMax, be...

lgl (1): is_train

Use `spec()` to retrieve the full column specification for this data.

Specify the column types or set `show_col_types = FALSE` to quiet this message.

2. Archivo de geometrías JSON

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if using this option

```
[7]: # file_data <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

```
[8]: iFileg <- "CU_04_05_01_zonasgeo.json"
file_geo <- paste0(iPath, iFileg)

if(file.exists(file_geo)){
  cat("Se leerán datos del archivo: ", file_geo)
} else{
  warning("Cuidado: el archivo no existe.")
}
```

Se leerán datos del archivo: Data/Input/CU_04_05_01_zonasgeo.json

Data file to dataframe Usar la función adecuada según el formato de entrada (xlsx, csv, json, ...)

```
[9]: datageo <- st_read(file_geo)
```

```
Reading layer `CU_04_05_01_zonasgeo' from data source
`/home/cegf/Workspace/CITY/CitizenLab-Research-and-Development/casos_urjc/note
books/dominios_II_y_III/04_vacunas/Data/Input/CU_04_05_01_zonasgeo.json'
using driver `GeoJSON'
Simple feature collection with 286 features and 3 fields
Geometry type: MULTIPOLYGON
Dimension: XY
Bounding box: xmin: -4.579396 ymin: 39.8848 xmax: -3.052977 ymax: 41.16584
Geodetic CRS: WGS 84
```

0.5 Data join

Unimos los dos data frames

```
[10]: dataj <- data |>
      full_join(datageo, by = c("GEOCODIGO", "DESBDT"))
```

0.6 Data Structure

Estructura de los datos:

```
[11]: dataj |> glimpse()
```

```

Rows: 21,740
Columns: 51
$ GEOCODIGO      <chr> "259", "260", "041", "025", "046",
"159", "065", "09...
$ DESBDT         <chr> "V Centenario", "Valdeacederas",
"Canillejas", "Bara...
$ ano            <dbl> 2022, 2022, 2022, 2022, 2022, 2022,
2022, 2021, 2023...
$ semana         <dbl> 34, 8, 9, 49, 24, 3, 8, 47, 1, 2,
52, 39, 16, 50, 34...
$ n_vacunas      <dbl> 0, 0, 0, 292, 0, 524, 0, 248, 204,
205, NA, 0, 0, 51...
$ n_citas        <dbl> 0, 0, 0, 280, 0, 498, 0, 228, 198,
187, NA, 0, 0, 51...
$ tmed          <dbl> 27.278748, 9.577289, 8.536554,
9.065363, 29.905728, ...
$ prec           <dbl> 0.169955881, 1.264910043,
3.122881160, 7.313886680, ...
$ velmedia       <dbl> 2.297067, 1.890425, 2.418071,
1.562328, 2.564749, 1...
$ presMax        <dbl> 940.0420, 944.1770, 949.7179,
941.8342, 940.5669, 95...
$ benzene        <dbl> 0.1764413, 0.4591543, 0.4099159,
0.4224172, 0.195865...
$ co             <dbl> 0.4987735, 0.3960647, 0.3951587,
NA, 0.2891224, 0.50...
$ no             <dbl> NA, 6.611337, 9.331224, 14.007722,
4.063517, 24.4756...
$ no2            <dbl> 14.21113, 34.67671, 30.29999,
32.54832, 26.06913, 44...
$ nox            <dbl> 18.00109, 48.94660, 45.22346,
56.75574, 30.35311, 74...
$ o3             <dbl> 80.90659, 42.06663, 48.88088,
26.68276, 64.55205, 31...
$ pm10           <dbl> 20.117087, 15.042152, 14.002432,
18.032354, 55.79346...
$ pm2.5          <dbl> 10.628064, 5.539590, 7.124192,
6.793868, 19.520373, ...
$ so2            <dbl> 2.794934, 3.507164, 2.692125,
2.351139, 3.397640, 2...
$ campana        <dbl> NA, NA, NA, 2022, NA, 2021, NA,
2021, 2022, 2021, 20...
$ scampana       <dbl> NA, NA, NA, 14, NA, 20, NA, 12, 18,
19, 17, 4, NA, 1...
$ capacidad_zona <dbl> 7957, 6537, 7167, 5633, 3864,
12583, 8544, 5077, 494...
$ prop_riesgo    <dbl> 0.11393237, 0.15763986, 0.25500690,
0.14452370, 0.26...

```

```

$ tasa_riesgo      <dbl> 0.013477754, 0.015731142,
0.009177382, 0.013099129, ...
$ tasa_mayores     <dbl> 0.023033610, 0.032817374,
0.028147027, 0.020829657, ...
$ poblacion_mayores <dbl> 0.10330662, 0.14362062, 0.23161874,
0.13058449, 0.24...
$ nombre_zona      <chr> "V Centenario", "Valdeacederas",
"Canillejas", "Bara...
$ nsec             <dbl> 17, 18, 22, 13, 14, 42, 32, 13, 17,
11, NA, 15, 15, ...
$ t3_1             <dbl> 36.73039, 41.41412, 45.44882,
39.78001, 46.13171, 46...
$ t1_1             <dbl> 31778, 26202, 28658, 22492, 15450,
50478, 34148, 202...
$ t2_1             <dbl> 0.5084658, 0.5329728, 0.5316594,
0.5189021, 0.551191...
$ t2_2             <dbl> 0.4915342, 0.4670272, 0.4683406,
0.4810979, 0.448809...
$ t4_1             <dbl> 0.22551283, 0.12790298, 0.12603707,
0.18104432, 0.11...
$ t4_2             <dbl> 0.6711962, 0.7284970, 0.6423306,
0.6883785, 0.641173...
$ t4_3             <dbl> 0.10330662, 0.14362062, 0.23161874,
0.13058449, 0.24...
$ t5_1             <dbl> 0.1063332, 0.2295250, 0.1655070,
0.1266086, 0.165893...
$ t6_1             <dbl> 0.1706875, 0.3477631, 0.2511757,
0.1998911, 0.261480...
$ t7_1             <dbl> 0.05131106, 0.04606911, 0.04379644,
0.05585777, 0.06...
$ t8_1             <dbl> 0.03892836, 0.03586418, 0.03207779,
0.04434976, 0.05...
$ t9_1             <dbl> 0.5151383, 0.3863876, 0.3129631,
0.4611972, 0.701812...
$ t10_1            <dbl> 0.09258503, 0.13151901, 0.13926119,
0.10460043, 0.06...
$ t11_1            <dbl> 0.6406787, 0.5451465, 0.4600730,
0.5920292, 0.471769...
$ t12_1            <dbl> 0.7028586, 0.6277335, 0.5346482,
0.6590530, 0.502531...
$ area             <dbl> 2100118.9, 1164622.0, 1597474.5,
3816572.0, 870986.8...
$ densidad_hab_km  <dbl> 15131.52443, 22498.28643,
17939.56640, 5893.24662, 1...
$ tuits_gripe      <dbl> 60, 56, 72, 196, 46, 382, 56, 280,
24, 508, NA, 126,...
$ interes_gripe    <dbl> 24, 15, 24, 77, 21, 42, 15, 64, 64,
69, NA, 42, 40, ...

```

```
$ Target          <dbl> 0, 0, 0, 292, 0, 524, 0, 248, 204,
205, NA, 0, 0, 51...
$ is_train        <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE,
TRUE, TRUE, TRUE...
$ CODBDT          <int> 686471, 686472, 686253, 686237,
686258, 686371, 6862...
$ geometry        <MULTIPOLYGON [°]> MULTIPOLYGON
(((−3.613203 4..., MULTIPO...
```

Muestra de los primeros datos:

```
[12]: dataj |> slice_head(n = 5)
```

	GEOCODIGO <chr>	DESBDT <chr>	ano <dbl>	semana <dbl>	n_vacunas <dbl>	n_citas <dbl>	tmed <dbl>	
	259	V Centenario	2022	34	0	0	27.278748	0
A spec_tbl_df: 5 × 51	260	Valdeacederas	2022	8	0	0	9.577289	1
	041	Canillejas	2022	9	0	0	8.536554	3
	025	Barajas	2022	49	292	280	9.065363	7
	046	Castelló	2022	24	0	0	29.905728	0

0.7 Data Visualization

0.7.1 Map

Filtrar y agrupar los datos a mostrar en el mapa

```
[13]: mdata <- dataj |>
      group_by(geometry) |>
      summarise(valor = sum(Target, na.rm = TRUE))
```

```
[14]: pal <- colorNumeric(palette = "Blues",
                        domain = mdata$valor)

sf::st_as_sf(mdata) |>
  leaflet() |>
  addTiles() |>
  addPolygons(color = "#444444",
              weight = 1,
              smoothFactor = 0.5,
              fillOpacity = 1,
              fillColor = ~pal(valor),
              highlightOptions = highlightOptions(color = "white", weight = 2,
                                                    bringToFront = TRUE),
              # popup = ~paste0(DESBDT, " (", GEOCODIGO, ")"),
              label = ~paste0(valor, " vacunas")) |>
  addLegend("bottomright",
           pal = pal,
           values = ~valor,
```

```
        title = "Número de vacunas",  
        labFormat = labelFormat(big.mark = " " ),  
        opacity = 1  
    )
```

HTML widgets cannot be represented in plain text (need html)

[]:

0.8 REPORT

A continuación se realizará un informe de las acciones realizadas

0.9 Main Actions Carried Out

- Se han realizado visualizaciones adicionales de los datos del caso de uso

0.10 Main Conclusions

- Los datos son adecuados para el caso de uso