

## 05. - Data Collection\_CU\_18\_02\_infracsv\_v\_01

June 13, 2023

#

CU18\_Infraestructuras\_eventos

Citizenlab Data Science Methodology > II - Data Processing Domain \*\*\* > # 05.- Data Collection

Data Collection is the process to obtain and generate (if required) necessary data to model the problem.

### 0.0.1 02. Crear csv con datos de infraestructuras

- Consolidar los datos en formato shp en data.frames con coordenadas
- Unir datos de hospitales de cu\_25
- Total 12 ficheros de entrada y uno de salida

Table of Contents

Settings

Data Load

ETL Processes

Import data from: CSV, Excel, Tab, JSON, SQL, and Parquet files

Synthetic Data Generation

Fake Data Generation

Open Data

Data Save

Main Conclusions

Main Actions

Acciones done

Acctions to perform

## 0.1 Settings

### 0.1.1 Packages to use

- {tcltk} para selección interactiva de archivos locales
- {sf} para datos georeferenciados

- {dplyr} para exploración y manipulación de datos
- {readr} para leer y escribir archivos csv

```
[1]: library(sf)
      library(dplyr)
      library(readr)
```

Linking to GEOS 3.10.2, GDAL 3.4.2, PROJ 8.2.1; sf\_use\_s2() is TRUE

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

### 0.1.2 Paths

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

## 0.2 Data Load

If there are more than one input file, make as many sections as files to import.

Instrucciones - Los ficheros de entrada del proceso están siempre en Data/Input/.

- Si hay más de un fichero de entrada, se crean tantos objetos iFile\_xx y file\_data\_xx como ficheros de entrada (xx número correlativo con dos dígitos, rellenar con ceros a la izquierda)

### 0.2.1 1. Intercambiadores

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[3]: # file_data_01 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[4]: iFile_01 <- "Transportes y Comunicaciones_ Intercambiadores/intercambiadores.
      ↪shp"
```

```
file_data_01 <- paste0(iPath, iFile_01)
```

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

```
[5]: data_01 <- st_read(file_data_01,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

```
options:          ENCODING=latin1
Reading layer `intercambiadores' from data source
  `/Users/emilio.lcano/academico/gh_repos/_transferencia/citizenlab/CitizenLab-
Research-and-Development/casos_urjc/notebooks/II_data_processing/18_infraestruct
uras/Data/Input/Transportes y Comunicaciones_
Intercambiadores/intercambiadores.shp'
  using driver `ESRI Shapefile'
Simple feature collection with 24 features and 8 fields
Geometry type: POINT
Dimension:      XY
Bounding box:   xmin: 434246 ymin: 4466946 xmax: 451488 ymax: 4482485
Projected CRS: ETRS89 / UTM zone 30N
```

Estructura de los datos:

```
[6]: glimpse(data_01)
```

```
Rows: 24
Columns: 9
$ CMUN      <chr> "079", "079", "079", "079", "079", "079",
"079", "079", "079..."
$ ETIQUETA  <chr> "Grandes Intercambiadores Plaza Elíptica",
"Grandes Intercam..."
$ LOCALIDAD <chr> "Madrid", "Madrid", "Madrid", "Madrid",
"Madrid", "Madrid", ...
$ DIRECCION <chr> "Plaza Elíptica s/n", "Calle Princesa, 89",
"Estación Princi..."
$ CODIGO2   <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, 1...
$ UTM_X     <dbl> 439178, 438974, 438957, 441586, 442605,
442151, 441253, 4424...
$ UTM_Y     <dbl> 4470780, 4476259, 4474712, 4479841,
4476589, 4480368, 447757...
$ COD_SUCA  <chr> "07902184", "07904939", "07905571",
"07905377", "07900312", ...
$ geometry  <POINT [m]> POINT (439178.2 4470780), POINT
(438974 4476259), POIN...
```

Muestra de datos:

```
[7]: data_01 |> slice_head(n = 5)
```

Registered S3 method overwritten by 'geojsonsf':

```
method      from
print.geojson geojson
```

	CMUN <chr>	ETIQUETA <chr>	LOCALIDAD <chr>	DIRECCION <chr>
A sf: 5 × 9	079	Grandes Intercambiadores Plaza Elíptica	Madrid	Plaza Elíptica s/n
	079	Grandes Intercambiadores Moncloa	Madrid	Calle Princesa, 89
	079	Grandes Intercambiadores Príncipe Pío	Madrid	Estación Príncipe Pío
	079	Grandes Intercambiadores Plaza de Castilla	Madrid	Calle San Aquilino y Ave
	079	Grandes Intercambiadores Avenida de América	Madrid	Avenida de América 9 A

## 0.2.2 2. Metro

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[8]: # file_data_02 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[9]: iFile_02 <- "Transportes y comunicaciones_ Metro (bocas)/"
file_data_02 <- paste0(iPath, iFile_02)
```

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

```
[10]: data_02 <- st_read(file_data_02,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

options: ENCODING=latin1

Reading layer `bocas' from data source

```
`/Users/emilio.lcano/academico/gh_repos/_transferencia/citizenlab/CitizenLab-
Research-and-Development/casos_urjc/notebooks/II_data_processing/18_infraestruct
uras/Data/Input/Transportes y comunicaciones_ Metro (bocas)'
```

```
using driver `ESRI Shapefile'
```

Simple feature collection with 771 features and 11 fields

Geometry type: POINT

Dimension: XY

Bounding box: xmin: 423308.6 ymin: 4459431 xmax: 461969.4 ymax: 4490120

Projected CRS: ETRS89 / UTM zone 30N

Estructura de los datos:

```
[11]: glimpse(data_02)
```

```
Rows: 771
Columns: 12
$ ESTACION   <chr> "Plaza de Castilla", "Plaza de Castilla",
"Plaza de Castill...
$ LINEAS     <chr> "1, 10, 9", "1, 10, 9", "1, 10, 9", "1",
"1", "1", "1", "1"...
$ SALIDA     <chr> "Castellana", "Plaza de Castilla", "Pº
Castellana, impares"...
$ DIRECCION  <chr> "Frente depósito C. Isabel II", "Paseo de
la Castellana, bu...
$ MUNICIPIO  <chr> "Madrid", "Madrid", "Madrid", "Madrid",
"Madrid", "Madrid",...
$ COD_MUNI   <chr> "079", "079", "079", "079", "079", "079",
"079", "079", "07...
$ HORARIO    <chr> "Abierto de 6:00 a 1:30, a excepción de
los accesos con hor...
$ MECANIZADO <chr> "Mecanizado: 6:00 a 13:50, S,D, Festivo y
Verano de 2/6 a 2...
$ UTM_X      <dbl> 441600, 441574, 441545, 441012, 441161,
440713, 440830, 440...
$ UTM_Y      <dbl> 4479652, 4479802, 4479841, 4479502,
4479577, 4479032, 44791...
$ CODIGO2    <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, ...
$ geometry   <POINT [m]> POINT (441599.6 4479652), POINT
(441573.6 4479802), P...
```

Muestra de datos:

```
[12]: data_02 |> slice_head(n = 5)
```

	ESTACION <chr>	LINEAS <chr>	SALIDA <chr>	DIRECCION <chr>	
A sf: 5 × 12	Plaza de Castilla	1, 10, 9	Castellana	Frente depósito C. Isabel II	M
	Plaza de Castilla	1, 10, 9	Plaza de Castilla	Paseo de la Castellana, bulevard central	M
	Plaza de Castilla	1, 10, 9	Pº Castellana, impares	Paseo Castellana, 189 (junto a torres)	M
	Valdeacederas	1	Capitán Blanco Argibay	Calle Bravo Murillo, 324	M
	Valdeacederas	1	Aníbal	Calle Bravo Murillo, 350	M

### 0.2.3 3. Aeropuertos

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[13]: # file_data_03 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[14]: iFile_03 <- "Transportes y comunicaciones_ Aeropuertos/"
      file_data_03 <- paste0(iPath, iFile_03)
```

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

```
[15]: data_03 <- st_read(file_data_03,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

```
options:          ENCODING=latin1
Reading layer `aeropuertos' from data source
`/Users/emilio.lcano/academico/gh_repos/_transferencia/citizenlab/CitizenLab-
Research-and-Development/casos_urjc/notebooks/II_data_processing/18_infraestruct
uras/Data/Input/Transportes y comunicaciones_ Aeropuertos'
  using driver `ESRI Shapefile'
Simple feature collection with 3 features and 7 fields
Geometry type: POINT
Dimension:      XY
Bounding box:   xmin: 434226.4 ymin: 4469153 xmax: 462119.5 ymax: 4480754
Projected CRS: ETRS89 / UTM zone 30N
```

Estructura de los datos:

```
[16]: glimpse(data_03)
```

```
Rows: 3
Columns: 8
$ CMUN      <chr> "079", "065", "148"
$ BUSCA     <chr> "Aeropuerto Adolfo Suárez-Madrid Barajas",
"Aeropuerto Madri...
$ DIRECCIÓN <chr> "Avda. Hispanidad, s/n", "Ctra. Barrio de
la Fortuna, s/n", ...
$ MUNICIPIO <chr> "MADRID", "MADRID", "TORREJÓN DE ARDOZ"
$ CODIGO2   <int> 1, 2, 3
$ ETIQUETA  <chr> "Adolfo Suárez-Madrid Barajas", "Madrid-
Cuatro Vientos", "Ma...
$ INFOR     <chr> "Madrid-Barajas<br>Avda. Hispanidad, s/n",
"Madrid-Cuatro Vi...
$ geometry  <POINT [m]> POINT (451504.3 4478985), POINT
(434226.4 4469153), POINT (4...
```

Muestra de datos:

```
[17]: data_03 |> slice_head(n = 5)
```

	CMUN <chr>	BUSCA <chr>	DIRECCIÓN <chr>	MUNICIPIO <chr>
A sf: 3 × 8	079	Aeropuerto Adolfo Suárez-Madrid Barajas	Avda. Hispanidad, s/n	MADRID
	065	Aeropuerto Madrid-Cuatro Vientos	Ctra. Barrio de la Fortuna, s/n	MADRID
	148	Aeropuerto Madrid-Torrejón de Ardoz	Ctra. de la Base, s/n	TORREJÓN

#### 0.2.4 4. Estaciones de Cercanías

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[18]: # file_data_04 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[19]: iFile_04 <- "Transportes y comunicaciones_ Cercanías (estaciones)/estcerca.shp"
file_data_04 <- paste0(iPath, iFile_04)
```

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

```
[20]: data_04 <- st_read(file_data_04,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

options: ENCODING=latin1

Reading layer `estcerca' from data source

`/Users/emilio.lcano/academico/gh\_repos/\_transferencia/citizenlab/CitizenLab-Research-and-Development/casos\_urjc/notebooks/II\_data\_processing/18\_infraestructuras/Data/Input/Transportes y Comunicaciones\_ Cercanías (estaciones)/estcerca.shp'

using driver `ESRI Shapefile'

Simple feature collection with 93 features and 8 fields

Geometry type: POINT

Dimension: XY

Bounding box: xmin: 900524.4 ymin: 4445522 xmax: 982989.2 ymax: 4531246

Projected CRS: ETRS89 / UTM zone 29N

Estructura de los datos:

```
[21]: glimpse(data_04)
```

Rows: 93

Columns: 9

\$ LINEA <chr> "C-2, C-7, C-8", "C-2, C-7, C-8", "C-2, C-7, C-8", "C-2, C-7...

\$ MUNICIPIO <chr> "Alcalá de Henares", "Torrejón de Ardoz", "Coslada", "Coslad...

```
$ CMUN      <chr> "005", "148", "049", "049", "079", "079",
"079", "079", "079...
$ CONEXION  <chr> "Aparcamiento", "Aparcamiento",
"Aparcamiento", "Aparcamient...
$ ETIQUETA  <chr> "Alcalá de Henares", "Torrejón de Ardoz",
"San Fernando de H...
$ UTM_X     <int> 468990, 459317, 454718, 452407, 449429,
448313, 446984, 4442...
$ UTM_Y     <int> 4482106, 4478341, 4477023, 4474944,
4472464, 4470879, 447034...
$ COD_SUCA  <chr> "005A0549", "14801866", "04910029",
"049A0172", "07908410", ...
$ geometry  <POINT [m]> POINT (977597.5 4497316), POINT
(968176.9 4492889), PO...
```

Muestra de datos:

```
[22]: data_04 |> slice_head(n = 5)
```

	LINEA <chr>	MUNICIPIO <chr>	CMUN <chr>	CONEXION <chr>	ETIQUETA <chr>
A sf: 5 × 9	C-2, C-7, C-8	Alcalá de Henares	005	Aparcamiento	Alcalá de Henare
	C-2, C-7, C-8	Torrejón de Ardoz	148	Aparcamiento	Torrejón de Ardo
	C-2, C-7, C-8	Coslada	049	Aparcamiento	San Fernando de
	C-2, C-7, C-8	Coslada	049	Aparcamiento, conexión L-7 metro	Coslada
	C-2, C-7, C-8	Madrid	079	Aparcamiento, conexión L-9 metro	Vicálvaro

## 0.2.5 5. Helisuperficies

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[23]: # file_data_05 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[24]: iFile_05 <- "Transportes y comunicaciones_ Helisuperficies/"
file_data_05 <- paste0(iPath, iFile_05)
```

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

```
[25]: data_05 <- st_read(file_data_05,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

options: ENCODING=latin1  
Reading layer `helisuper' from data source



```
`/Users/emilio.lcano/academico/gh_repos/_transferencia/citizenlab/CitizenLab-
Research-and-Development/casos_urjc/notebooks/II_data_processing/18_infraestruct
uras/Data/Input/Transportes y comunicaciones_Helisuperficies'
```

```
using driver `ESRI Shapefile'
Simple feature collection with 87 features and 8 fields
Geometry type: POINT
Dimension:      XY
Bounding box:   xmin: 373754 ymin: 4434452 xmax: 490376 ymax: 4550969
Projected CRS:  ETRS89 / UTM zone 30N
```

Estructura de los datos:

[26]: `glimpse(data_05)`

```
Rows: 87
Columns: 9
$ CMUN      <chr> "0133", "0604", "0528", "1570", "0552",
"0259", "0915", "091...
$ MUNICIPIO <chr> "Aranjuez", "Fuentidueña de Tajo",
"Chinchón", "Valdelaguna"...
$ ETIQUETA  <chr> "Helipuerto Hospital del Tajo", "Helipuerto
Eventual de Fuen...
$ USO       <chr> "Sanitarias/ Emergencias", "Emergencias
Sanitarias", "Emerge...
$ GESTOR    <chr> "Hospital del Tajo (Aranjuez)",
"Ayuntamiento de Fuentidueña...
$ UTM_X     <dbl> 447862, 485991, 463295, 467472, 490106,
490376, 461962, 4591...
$ UTM_Y     <dbl> 4434452, 4441209, 4444189, 4445339,
4449213, 4453110, 445350...
$ CLAVE     <chr> "HPUR", "HUEOEE", "HUEOEE", "ACUPUR",
"HUEOEE", "HUEOEE", "H...
$ geometry  <POINT [m]> POINT (447862 4434452), POINT (485991
4441209), POINT ...
```

Muestra de datos:

[27]: `data_05 |> slice_head(n = 5)`

	CMUN <chr>	MUNICIPIO <chr>	ETIQUETA <chr>	USO <chr>
	0133	Aranjuez	Helipuerto Hospital del Tajo	Sanitarias/ Emergen
A sf: 5 × 9	0604	Fuentidueña de Tajo	Helipuerto Eventual de Fuentidueña del Tajo	Emergencias Sanitar
	0528	Chinchón	Helipuerto Eventual de Chinchón	Emergencias Sanitar
	1570	Valdelaguna	Campo de Ultraligeros de Valdelaguna	Aviación Deportiva/
	0552	Estremera	Helipuerto Eventual de Estremera	Emergencias Sanitar

## 0.2.6 6. Centros de Atención Integral a Drogodependiente

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Uncomment the line if not using this option

```
[28]: # file_data_06 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[29]: iFile_06 <- "Salud_Centros sanitarios. Centros de Atención Integral a
↳Drogodependientes/drogodependencia.shp"
file_data_06 <- paste0(iPath, iFile_06)
```

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

```
[30]: data_06 <- st_read(file_data_06,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

options: ENCODING=latin1

Reading layer 'drogodependencia' from data source

`/Users/emilio.lcano/academico/gh\_repos/\_\_transferencia/citizenlab/CitizenLab-Research-and-Development/casos\_urjc/notebooks/II\_data\_processing/18\_infraestructuras/Data/Input/Salud\_Centros sanitarios. Centros de Atención Integral a Drogodependientes/drogodependencia.shp'

using driver 'ESRI Shapefile'

Simple feature collection with 35 features and 9 fields

Geometry type: POINT

Dimension: XY

Bounding box: xmin: 380860 ymin: 4431637 xmax: 468812 ymax: 4501031

Projected CRS: ETRS89 / UTM zone 30N

Estructura de los datos:

```
[31]: glimpse(data_06)
```

Rows: 35

Columns: 10

\$ CODMUN <chr> "005", "006", "007", "013", "014", "047",  
"045", "049", "079...

\$ NOMBRE <chr> "Centro de Atención Integral a  
Drogodependientes Alcalá de H...

\$ DIRECCION <chr> "C/ Santiago, 13 - planta", "C/ Libertad, 6  
- 2º", "Paseo de...

\$ UTM\_X <dbl> 468812, 446255, 429345, 449413, 463258,  
414669, 434697, 4531...

\$ UTM\_Y <dbl> 4481428, 4487944, 4466846, 4431637,  
4460592, 4498050, 450103...

\$ MUNICIPIO <chr> "Alcalá de Henares", "Alcobendas",  
"Alcorcón", "Aranjuez", "..."

```
$ URL      <chr>
"http://centrossanitarios.sanidadmadrid.org/Otros/Detalle.as...
$ CODIGO2   <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, 1...
$ COD_SUCA  <chr> "005A0696", "006A0179", "00701867",
"013A0208", "014A0253", ...
$ geometry  <POINT [m]> POINT (468812 4481428), POINT (446255
4487944), POINT ...
```

Muestra de datos:

```
[32]: data_06 |> slice_head(n = 5)
```

	CODMUN <chr>	NOMBRE <chr>	DIRECCION <chr>
	005	Centro de Atención Integral a Drogodependientes Alcalá de Henares	C/ Santiago, 13
A sf: 5 × 10	006	Centro de Atención Integral a Drogodependientes Alcobendas	C/ Libertad, 6
	007	Centro de Atención Integral a Drogodependientes Alcorcón	Paseo de Extremadura
	013	Centro de Atención Integral a Drogodependientes Aranjuez	C/ Primero de Mayo
	014	Centro de Atención Integral a Drogodependientes Arganda del Rey	C/ Camino del Rey

## 0.2.7 7. Centros de especialidades

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[33]: # file_data_07 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[34]: iFile_07 <- "Salud_ Centros sanitarios. Centros de especialidades/
↪centroespecialidades.shp"
file_data_07 <- paste0(iPath, iFile_07)
```

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

```
[35]: data_07 <- st_read(file_data_07,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

options: ENCODING=latin1

Reading layer 'centroespecialidades' from data source

```
`/Users/emilio.lcano/academico/gh_repos/_transferencia/citizenlab/CitizenLab-
Research-and-Development/casos_urjc/notebooks/II_data_processing/18_infraestruct
uras/Data/Input/Salud_ Centros sanitarios. Centros de
especialidades/centroespecialidades.shp'
using driver 'ESRI Shapefile'
```

Simple feature collection with 28 features and 15 fields  
Geometry type: POINT  
Dimension: XY  
Bounding box: xmin: 402704.8 ymin: 4454857 xmax: 468379.8 ymax: 4501383  
Projected CRS: ETRS89 / UTM zone 30N

Estructura de los datos:

```
[36]: glimpse(data_07)
```

```
Rows: 28
Columns: 16
$ ETIQUETA <chr> "C.E. Moratalaz", "C.E. Peña Prieta -
Hermanos Sangr ", "C.E...
$ TIPOCEN2 <chr> "Centro de Especialidades", "Centro de
Especialidades", "Cen...
$ TIPOVIAL <chr> "Calle", "Avenida", "Calle", "Avenida",
"Calle", "Avenida", ...
$ NOMVIA <chr> "Hacienda de Pavones", "Pe a Prieta",
"Sierra de Alquife", "...
$ NUMERO <chr> "348", "4", "8", "285", "45", "s/n", "11",
"30", "89", "13",...
$ MUNICIPIO <chr> "Madrid", "Madrid", "Madrid", "Madrid",
"Madrid", "Torrej n ...
$ CODMUN <chr> "0796", "0796", "0796", "0796", "0796",
"1489", "0053", "079...
$ CODPOS <chr> "28030", "28038", "28018", "28038",
"28028", "28850", "28806...
$ CODAREA <chr> "01", "01", "01", "01", "02", "03", "03",
"04", "04", "05", ...
$ CODDISTR <chr> "0102", "0104", "0104", "0104", "0202",
"0302", "0301", "040...
$ CODZBS <chr> "010203", "010402", "010406", "010410",
"020201", "030202", ...
$ CODVIA <chr> "07904590", "07904676", "07909818",
"07900130", "07901736", ...
$ UTM_X <dbl> 446223, 443279, 443380, 445704, 443300,
459896, 468380, 4460...
$ UTM_Y <dbl> 4472348, 4472215, 4471141, 4470967,
4474985, 4479414, 448330...
$ CODIGO2 <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14,
15, 16, 17, 18, 1...
$ geometry <POINT [m]> POINT (446222.8 4472348), POINT
(443278.6 4472215), PO...
```

Muestra de datos:

```
[37]: data_07 |> slice_head(n = 5)
```

	ETIQUETA <chr>	TIPOCEN2 <chr>	TIPOVIAL <chr>	NOMVIA <chr>
A sf: 5 × 16	C.E. Moratalaz	Centro de Especialidades	Calle	Hacienda de Pavon
	C.E. Peña Prieta - Hermanos Sangr�	Centro de Especialidades	Avenida	Pe�a Prieta
	C.E. Vicente Soldevilla	Centro de Especialidades	Calle	Sierra de Alquife
	C.E. Federica Montseny	Centro de Especialidades	Avenida	Albufera
	C.E. Hermanos Garc�a Noblejas	Centro de Especialidades	Calle	Doctor Esquerdo

### 0.2.8 8. Centros de salud

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[38]: # file_data_08 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresi n del c digo

```
[39]: iFile_08 <- "Salud_Centros sanitarios. Centros de salud/centrosalud.shp"
file_data_08 <- paste0(iPath, iFile_08)
```

**Data file to dataframe** Usar la funci n adecuada seg n el formato de entrada (xlsx, csv, json, ...)

```
[40]: data_08 <- st_read(file_data_08,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

options: ENCODING=latin1

Reading layer `centrosalud' from data source

```
`/Users/emilio.lcano/academico/gh_repos/_transferencia/citizenlab/CitizenLab-
Research-and-Development/casos_urjc/notebooks/II_data_processing/18_infraestruct
uras/Data/Input/Salud_Centros sanitarios. Centros de salud/centrosalud.shp'
using driver `ESRI Shapefile'
```

Simple feature collection with 267 features and 14 fields

Geometry type: POINT

Dimension: XY

Bounding box: xmin: 377838.2 ymin: 4430802 xmax: 476562 ymax: 4538179

Projected CRS: ETRS89 / UTM zone 30N

Estructura de los datos:

```
[41]: glimpse(data_08)
```

Rows: 267

Columns: 15

```
$ TIPOCEN2 <chr> "Centro de Salud Carmen Calzado", "Centro
de Salud Juan de ...
```

```
$ TIPOVIAL <chr> "Calle", "Avenida", "Calle", "Calle",
```

```

"Calle", "Calle", "Av...
$ NOMVIA      <chr> "Carmen Calzado", "Juan de Austria",
"Arturo Soria", "Luis ...
$ NUMERO      <chr> "14", "19", "12", "16", "1", "1", "23",
"12", "61", "12", "...
$ MUNICIPIO   <chr> "Alcalá de Henares", "Alcalá de Henares",
"Alcalá de Henare...
$ CMUN        <chr> "005", "005", "005", "005", "005", "005",
"005", "005", "00...
$ CODVIA      <chr> "005A0150", "005A0370", "00500730",
"005A0424", "005A0441",...
$ UTM_X       <dbl> 468979, 470297, 466617, 468796, 469712,
469685, 468549, 468...
$ UTM_Y       <dbl> 4481282, 4481952, 4481245, 4480551,
4481336, 4482126, 44844...
$ CODIGO2     <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, ...
$ SUCA_CODVI  <chr> "005A0150", "005A0370", "00500730",
"005A0424", "005A0441",...
$ CODZBS      <chr> "030101", "030107", "030109", "030104",
"030105", "030108",...
$ CODDISTR    <chr> "0301", "0301", "0301", "0301", "0301",
"0301", "0301", "03...
$ CODAREA     <chr> "03", "03", "03", "03", "03", "03", "03",
"03", "03", "03",...
$ geometry    <POINT [m]> POINT (468979 4481282), POINT
(470297 4481952), POINT...

```

Muestra de datos:

```
[42]: data_08 |> slice_head(n = 5)
```

	TIPOCEN2 <chr>	TIPOVIAL <chr>	NOMVIA <chr>	NUMERO <chr>	MUNICIPIO <chr>
A sf: 5 × 15	Centro de Salud Carmen Calzado	Calle	Carmen Calzado	14	Alcalá de Henares
	Centro de Salud Juan de Austria	Avenida	Juan de Austria	19	Alcalá de Henares
	Centro de Salud La Garena	Calle	Arturo Soria	12	Alcalá de Henares
	Centro de Salud Luis Vives	Calle	Luis Vives	16	Alcalá de Henares
	Centro de Salud Manuel Merino	Calle	Manuel Merino	1	Alcalá de Henares

### 0.2.9 9. Consultorios de salud

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[43]: # file_data_09 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[44]: iFile_09 <- "Salud_Centros sanitarios. Consultorios de salud/consultoriosalud.
      ↪shp"
      file_data_09 <- paste0(iPath, iFile_09)
```

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

```
[45]: data_09 <- st_read(file_data_09,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

```
options:          ENCODING=latin1
Reading layer `consultoriosalud' from data source
  `~/Users/emilio.lcano/academico/gh_repos/_transferencia/citizenlab/CitizenLab-
Research-and-Development/casos_urjc/notebooks/II_data_processing/18_infraestruct
uras/Data/Input/Salud_Centros sanitarios. Consultorios de
salud/consultoriosalud.shp'
  using driver `ESRI Shapefile'
Simple feature collection with 156 features and 10 fields
Geometry type: POINT
Dimension:      XY
Bounding box:   xmin: 373181 ymin: 4435469 xmax: 490638 ymax: 4553677
Projected CRS: ETRS89 / UTM zone 30N
```

Estructura de los datos:

```
[46]: glimpse(data_09)
```

```
Rows: 156
Columns: 11
$ ETIQUETA <chr> "Consultorio de La Acebeda", "Consultorio
de Ajalvir", "Cons...
$ TIPOVIA <chr> "Plaza", "Avda", "Plaza", "Calle", "Calle",
"Avda", "Plaza",...
$ NOMVIA <chr> "San Miguel", "Aguas", "Santa Marina",
"Federico García Lorc...
$ NUMERO <chr> "1", "2", "17", "31", "17", "3", "1", "15",
"2", "11", "s/n"...
$ MUNICIPIO <chr> "La Acebeda", "Ajálvir", "Alameda del
Valle", "El Álamo", "A...
$ CODMUNI <chr> "0014", "0029", "0035", "0040", "0088",
"0091", "0105", "011...
$ CODVIA <chr> "00100028", "00200141", "003A0030",
"00400218", "008A0088", ...
$ UTM_X <int> 447522, 459130, 428958, 415892, 398120,
450987, 413034, 4844...
$ UTM_Y <dbl> 4548589, 4487429, 4530000, 4454069,
4464011, 4496981, 450118...
$ COD2 <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
```

```
14, 15, 16, 17, 1...
$ geometry <POINT [m]> POINT (447522 4548589), POINT (459130
4487429), POINT ...
```

Muestra de datos:

```
[47]: data_09 |> slice_head(n = 5)
```

	ETIQUETA <chr>	TIPOVIA <chr>	NOMVIA <chr>	NUMERO <chr>	MUNICIPIO <chr>
A sf: 5 × 11	Consultorio de La Acebeda	Plaza	San Miguel	1	La Acebeda
	Consultorio de Ajalvir	Avda	Aguas	2	Ajalvir
	Consultorio de Alameda del Valle	Plaza	Santa Marina	17	Alameda del Valle
	Consultorio de El Álamo	Calle	Federico García Lorca	31	El Álamo
	Consultorio de Aldea del Fresno	Calle	San Rafael	17	Aldea del Fresno

## 0.2.10 10. Otros Centros

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[48]: # file_data_10 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[49]: iFile_10 <- "Salud_Centros sanitarios. Otros centros/saludotros.shp"
file_data_10 <- paste0(iPath, iFile_10)
```

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

```
[50]: data_10 <- st_read(file_data_10,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

options: ENCODING=latin1

Reading layer `saludotros' from data source

```
`/Users/emilio.lcano/academico/gh_repos/_transferencia/citizenlab/CitizenLab-
Research-and-Development/casos_urjc/notebooks/II_data_processing/18_infraestruct
uras/Data/Input/Salud_Centros sanitarios. Otros centros/saludotros.shp'
```

using driver `ESRI Shapefile'

Simple feature collection with 29 features and 7 fields

Geometry type: POINT

Dimension: XY

Bounding box: xmin: 434198.7 ymin: 4463663 xmax: 468813 ymax: 4489946

Projected CRS: ETRS89 / UTM zone 30N

Estructura de los datos:



```
[51]: glimpse(data_10)
```

```
Rows: 29
Columns: 8
$ CODMUN    <chr> "079", "079", "079", "079", "079", "079",
"134", "079", "079...
$ NOMBRE    <chr> "Instituto Provincial de Rehabilitación",
"Instituto Provinc...
$ DIRECCION <chr> "Calle Francisco Silvela, 40", "Calle
General Arrando, 17", ...
$ CODIGO2   <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, 1...
$ UTM_X     <dbl> 442944, 441216, 441846, 441859, 441873,
441821, 448297, 4477...
$ UTM_Y     <dbl> 4476050, 4475926, 4480831, 4480830,
4480830, 4476418, 448994...
$ COD_SUCA  <chr> "07902340", "07902499", "07901213",
"07901213", "07901213", ...
$ geometry  <POINT [m]> POINT (442944 4476050), POINT (441216
4475926), POINT ...
```

Muestra de datos:

```
[52]: data_10 |> slice_head(n = 5)
```

	CODMUN <chr>	NOMBRE <chr>	DIRECCION <chr>	COD <int>
A sf: 5 × 8	079	Instituto Provincial de Rehabilitación	Calle Francisco Silvela, 40	1
	079	Instituto Provincial de Oftalmología	Calle General Arrando, 17	2
	079	Oficina Regional de Coordinación de Transplantes	Paseo de la Castellana, 280	3
	079	Oficina Regional de Coordinación de Salud Mental	Paseo de la Castellana, 280	4
	079	Oficina Regional de Coordinación Oncológica	Paseo de la Castellana, 280	5

### 0.2.11 11. Centros de salud mental

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[53]: # file_data_11 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[54]: iFile_11 <- "Salud_ Centros sanitarios. Servicio de salud mental de distrito/
↪saludmental.shp"
file_data_11 <- paste0(iPath, iFile_11)
```

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

```
[55]: data_11 <- st_read(file_data_11,
                        options = "ENCODING=latin1",
                        quiet = FALSE)
```

```
options:          ENCODING=latin1
Reading layer `saludmental' from data source
  `/Users/emilio.lcano/academico/gh_repos/__transferencia/citizenlab/CitizenLab-
Research-and-Development/casos_urjc/notebooks/II_data_processing/18_infraestruct
uras/Data/Input/Salud_Centros sanitarios. Servicio de salud mental de
distrito/saludmental.shp'
  using driver `ESRI Shapefile'
Simple feature collection with 53 features and 10 fields
Geometry type: POINT
Dimension:      XY
Bounding box:   xmin: 402720 ymin: 4434284 xmax: 468396 ymax: 4501363
Projected CRS: ETRS89 / UTM zone 30N
```

Estructura de los datos:

```
[56]: glimpse(data_11)
```

```
Rows: 53
Columns: 11
$ NOMBRE      <chr> "CSM Villa Vallecas", "CSM Coslada", "CSM
Salamanca", "CSM ...
$ TIPOVIAL    <chr> "Calle", "Avenida", "Calle", "Calle",
"Calle", "Avenida", "...
$ NOMVIA      <chr> "San Claudio", "España", "O'Donnell",
"Aguileñas", "Octavio...
$ MUNICIPIO   <chr> "Madrid", "Coslada", "Madrid", "Madrid",
"Alcalá de Henares...
$ NUMERO      <chr> "154", "s/n", "55", "1", "15", "4", "50",
"20", "1", "11", ...
$ CODMUN      <chr> "0796", "0493", "0796", "0796", "0053",
"1489", "0066", "13...
$ CODIGO2     <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14, 15, 16, 17, ...
$ SUCA_CDVIA  <chr> "07910251", "049A0101", "07904389",
"07900067", "00504560",...
$ UTM_X       <dbl> 445763, 453151, 443046, 441192, 468396,
461608, 446048, 446...
$ UTM_Y       <dbl> 4471085, 4474618, 4474762, 4479657,
4483323, 4478644, 44887...
$ geometry    <POINT [m]> POINT (445763 4471085), POINT
(453151 4474618), POINT...
```

Muestra de datos:

```
[57]: data_11 |> slice_head(n = 5)
```

	NOMBRE <chr>	TIPOVIAL <chr>	NOMVIA <chr>	MUNICIPIO <chr>	NUMERO <chr>	CO <chr>
A sf: 5 × 11	CSM Villa Vallecas	Calle	San Claudio	Madrid	154	0796
	CSM Coslada	Avenida	España	Coslada	s/n	0493
	CSM Salamanca	Calle	O'Donnell	Madrid	55	0796
	CSM Tetuán Infanto Juvenil	Calle	Aguileñas	Madrid	1	0796
	CSM Alcalá de Henarés I y II	Calle	Octavio Paz	Alcalá de Henares	15	0053

### 0.2.12 12. Hospitales

OPCION A: Seleccionar fichero en ventana para mayor comodidad

Data load using the {tcltk} package. Ucomment the line if not using this option

```
[58]: # file_data_12 <- tcltk::tk_choose.files(multi = FALSE)
```

OPCION B: Especificar el nombre de archivo

NOTA: se omiten las comprobaciones de existencia de archivo para mejor compresión del código

```
[59]: iFile_12 <- "hospitales_poi.csv"
file_data_12 <- paste0(iPath, iFile_12)
```

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

```
[60]: data_12 <- read_csv(file_data_12)
```

Rows: 87 Columns: 67

Column specification

Delimiter: ","

chr (31): Nombre Centro, Dirección, Teléfono, Municipio, Provincia, CCAA, Cód...

dbl (35): CCN, CODCNH, Cód. Municipio, Cód. Provincia, Cód. CCAA, Código Pos...

lgl (1): CIERREFECH

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.

Estructura de los datos:

```
[61]: glimpse(data_12)
```

Rows: 87

Columns: 67

\$ CCN <dbl> 1328020932, 1328000034, 1328000070, 1328000066...

\$ CODCNH <dbl> 281371, 280745, 281124,  
 281071, 281292, 281281...  
 \$ `Nombre Centro` <chr> "HOSPITAL HM VALLES",  
 "HOSPITAL UNIVERSITARIO ...  
 \$ Dirección <chr> "CALLE SANTIAGO 14",  
 "CARRETERA DE MECO S/N", ...  
 \$ Teléfono <chr> "918883262", "918878100",  
 "917810682", "916219...  
 \$ `Cód. Municipio` <dbl> 280053, 280053, 280072,  
 280072, 280133, 280148...  
 \$ Municipio <chr> "Alcalá de Henares", "Alcalá  
 de Henares", "Alc...  
 \$ `Cód. Provincia` <dbl> 28, 28, 28, 28, 28, 28, 28,  
 28, 28, 28, 28, 28...  
 \$ Provincia <chr> "MADRID", "MADRID", "MADRID",  
 "MADRID", "MADRI...  
 \$ `Cód. CCAA` <dbl> 13, 13, 13, 13, 13, 13, 13,  
 13, 13, 13, 13, 13...  
 \$ CCAA <chr> "MADRID", "MADRID", "MADRID",  
 "MADRID", "MADRI...  
 \$ `Código Postal` <dbl> 28801, 28805, 28922, 28922,  
 28300, 28500, 2866...  
 \$ CAMAS <dbl> 20, 507, 81, 402, 98, 132,  
 170, 53, 192, 339, ...  
 \$ `Cód. Clase de Centro` <chr> "C190", "C11", "C11", "C11",  
 "C11", "C11", "C1...  
 \$ `Clase de Centro` <chr> "Otros Centros con  
 Internamiento", "Hospitales...  
 \$ `Cód. Dep. Funcional` <dbl> 20, 2, 20, 2, 2, 2, 20, 20,  
 2, 20, 20, 2, 21, ...  
 \$ `Dependencia Funcional` <chr> "PRIVADOS", "SERVICIOS E  
 INSTITUTOS DE SALUD D...  
 \$ `Forma parte Complejo` <chr> "N", "N", "N", "N", "N", "N",  
 "N", "N", "N", "N", "...  
 \$ CODIDCOM <dbl> NA, NA, NA, NA, NA, NA, NA,  
 NA, NA, NA, NA, NA...  
 \$ `Nombre del Complejo` <chr> NA, NA, NA, NA, NA, NA, NA,  
 NA, NA, NA, NA, NA...  
 \$ ALTA <chr> "N", "N", "N", "N", "N", "N",  
 "N", "N", "N", "N", "...  
 \$ Email <chr> NA, NA, NA,  
 "FHALCORCON@FHALCORCON.ES", NA, NA...  
 \$ OBJECTID <dbl> 559, 516, 536, 532, 551, 550,  
 534, 541, 517, 5...  
 \$ NOMBRE <chr> "HM VALLES", "HOSPITAL  
 UNIVERSITARIO PRINCIPE ...  
 \$ DIRECCION <chr> "C/SANTIAGO, 14", "CTRA.  
 ALCALÁ-MECO, S/N", "E...

\$ TELEFONO <dbl> 917374353, 918878100,  
 916496600, 916219400, 91...  
 \$ TELEFONO2 <dbl> NA, NA, NA, NA, NA,  
 918394000, NA, NA, NA, 646...  
 \$ TELEFAX <dbl> 912805932, 918801825,  
 916496604, 916219902, 91...  
 \$ CODMU <dbl> 280053, 280053, 280072,  
 280072, 280133, 280148...  
 \$ MUNICIPIOS <chr> "Alcalá de Henares", "Alcalá  
 de Henares", "Alc...  
 \$ CODPROV <dbl> 28, 28, 28, 28, 28, 28, 28,  
 28, 28, 28, 28, 28...  
 \$ PROVINCIAS <chr> "MADRID", "MADRID", "MADRID",  
 "MADRID", "MADRI...  
 \$ CODAUTO <dbl> 13, 13, 13, 13, 13, 13, 13,  
 13, 13, 13, 13, 13...  
 \$ COMUNIDADES <chr> "MADRID", "MADRID", "MADRID",  
 "MADRID", "MADRI...  
 \$ CODPOSTAL <dbl> 28801, 28805, 28922, 28922,  
 28300, 28500, 2866...  
 \$ CODFI <dbl> 16, 1, 12, 1, 1, 1, 1, 1, 13,  
 6, 6, 1, 10, 1, ...  
 \$ FINALIDAD\_ASISITENCIAL <chr> "OTRA FINALIDAD", "GENERAL",  
 "MÉDICO-QUIRÚRGIC...  
 \$ CODPAT <dbl> 14, 1, 14, 1, 6, 6, 14, 14,  
 1, 12, 12, 6, 10, ...  
 \$ DEPENDENCIA\_PATRIMONIAL <chr> "PRIVADO NO BENÉFICO",  
 "SEGURIDAD SOCIAL", "PR...  
 \$ CODFU <dbl> 23, 34, 23, 34, 34, 34, 23,  
 23, 34, 21, 21, 34...  
 \$ DEPENDENCIA\_FUNCIONAL <chr> "PRIVADO NO BENÉFICO",  
 "SERVICIO MADRILEÑO DE ...  
 \$ ACREDOCENT <chr> "S", "S", "N", "S", "N", "N",  
 "S", "N", "S", "...  
 \$ ESCOMPLE <chr> "N", "N", "N", "N", "N", "N",  
 "N", "N", "N", "...  
 \$ FORCOMPLE <chr> "N", "N", "N", "N", "N", "N",  
 "N", "N", "N", "...  
 \$ CERRADO <chr> "N", "N", "N", "N", "N", "N",  
 "N", "N", "N", "...  
 \$ CAPITAL <chr> "N", "N", "N", "N", "N", "N",  
 "N", "N", "N", "...  
 \$ CIERREFECH <lg1> NA, NA, NA, NA, NA, NA, NA,  
 NA, NA, NA, NA, NA...  
 \$ CONCIERTO <chr> "S", "N", "S", "N", "N", "N",  
 "S", "S", "N", "...  
 \$ EMAIL <chr> "fgcalleja@hnhospitales.com;  
 imarco@hnhospital...

```

$ TAC <dbl> 1, 3, 1, 2, 1, 1, 1, 1, 0, 0,
0, 1, 1, 1, 1, 2...
$ RM <dbl> 1, 2, 2, 2, 1, 1, 3, 2, 0, 0,
0, 2, 2, 1, 2, 2...
$ GAM <dbl> 1, 1, 0, 1, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 1, 0...
$ HEM <dbl> 0, 1, 1, 1, 0, 0, 2, 0, 0, 0,
0, 1, 0, 0, 1, 1...
$ ASD <dbl> 0, 1, 1, 1, 1, 0, 2, 0, 0, 0,
0, 0, 0, 0, 1, 1...
$ LIT <dbl> 0, 0, 0, 1, 0, 0, 1, 0, 0, 0,
0, 0, 0, 0, 0, 0...
$ BCO <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0...
$ ALI <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 2, 0...
$ SPECT <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 1, 0, 0, 1, 2...
$ PET <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 1, 1...
$ MAMOS <dbl> 1, 2, 1, 2, 1, 1, 1, 1, 0, 0,
0, 1, 0, 2, 1, 1...
$ DO <dbl> 1, 1, 1, 1, 0, 0, 1, 0, 0, 0,
0, 1, 0, 0, 1, 0...
$ DIAL <dbl> 0, 28, 0, 28, 13, 18, 1, 0,
0, 0, 0, 18, 0, 25...
$ X <dbl> -3.367372, -3.347602,
-3.849035, -3.837394, -3...
$ Y <dbl> 40.48302, 40.51025, 40.34248,
40.34887, 40.057...
$ CalidadGeocodificacion <chr> "PointAddress", "Manual",
"StreetAddress", "Ma...
$ id_area <chr> "03", "03", "08", "08", "11",
"01", "06", "06"...
$ nombre_area <chr> "Este", "Este", "Sur-Oeste
I", "Sur-Oeste I", ...

```

Muestra de datos:

```
[62]: data_12 |> slice_head(n = 5)
```

	CCN <dbl>	CODCNH <dbl>	Nombre Centro <chr>
A spec_tbl_df: 5 × 67	1328020932	281371	HOSPITAL HM VALLES
	1328000034	280745	HOSPITAL UNIVERSITARIO PRINCIPE DE ASTURIAS
	1328000070	281124	HOSPITAL QUIRONSAUD SUR
	1328000066	281071	HOSPITAL UNIVERSITARIO FUNDACION ALCORCON
	1328000084	281292	HOSPITAL DEL TAJO S.A.

### 0.3 ETL Processes

#### 0.3.1 Import data from: CSV, Excel, Tab, JSON, SQL, and Parquet files

Se han importado en el apartado Data Load anterior:

- Localizaciones de POIS publicados en el portal de la Comunidad de Madrid
- Localización de hospitales del catálogo nacional de hospitales

Incluir apartados si procede para: Extracción de datos (select, filter), Transformación de datos, (mutate, joins, ...). Si es necesario tratar datos perdidos, indicarlo también en NB 09.2

#### 0.3.2 Data transformation

Para cada data frame de infraestructuras se realizan las siguientes transformaciones:

- Transformar sistema de referencia CRS a 4326 (WGS 84)
- Añadir las coordenadas como columnas normales
- Establecer variables **grupo** y **tipo**
- Si hay información adicional en el archivo, añadirla a columna **info**
- Asignar a la columna **nombre** la columna más adecuada según lo importado
- Seleccionar columnas que se usarán en todos los data frames a unir
- Eliminar la columna de geometría para que se pueda exportar a csv

Para los hospitales igual pero sin añadir coordenadas que ya las tenía ni eliminar la geometría porque no es objeto sf.

Finalmente se unen todos los data.frames en uno

```
[63]: tdata_01 <- data_01 |>
      bind_cols(st_coordinates(data_01 |> st_transform(4326))) |>
      mutate(grupo = "Transporte",
             tipo = "Intercambiadores",
             info = "") |>
      select(grupo,
             tipo,
             nombre = ETIQUETA,
             CODMUN = CMUN,
             DIRECCION,
             X, Y,
             info) |>
      st_drop_geometry()
```

```
[64]: tdata_01 |> slice_head(n = 5)
```

	grupo	tipo	nombre	CODMUN
	<chr>	<chr>	<chr>	<chr>
A data.frame: 5 × 8	Transporte	Intercambiadores	Grandes Intercambiadores Plaza Elíptica	079
	Transporte	Intercambiadores	Grandes Intercambiadores Moncloa	079
	Transporte	Intercambiadores	Grandes Intercambiadores Príncipe Pío	079
	Transporte	Intercambiadores	Grandes Intercambiadores Plaza de Castilla	079
	Transporte	Intercambiadores	Grandes Intercambiadores Avenida de América	079

```
[65]: tdata_02 <- data_02 |>
      bind_cols(st_coordinates(data_02 |> st_transform(4326))) |>
      mutate(grupo = "Transporte",
             tipo = "Bocas de metro",
             ESTACION = paste(ESTACION, SALIDA, sep = " - "),
             info = paste0("Líneas: ", LINEAS )) |>
      select(grupo,
             tipo,
             nombre = ESTACION,
             CODMUN = COD_MUNI,
             DIRECCION,
             X, Y,
             info) |>
      st_drop_geometry()
```

```
[66]: tdata_02 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	CODMUN <chr>	DIRECCION <chr>
A data.frame: 5 × 8	Transporte	Bocas de metro	Plaza de Castilla - Castellana	079	Fre
	Transporte	Bocas de metro	Plaza de Castilla - Plaza de Castilla	079	Pa
	Transporte	Bocas de metro	Plaza de Castilla - Pº Castellana, impares	079	Pa
	Transporte	Bocas de metro	Valdeacederas - Capitán Blanco Argibay	079	Ca
	Transporte	Bocas de metro	Valdeacederas - Aníbal	079	Ca

```
[67]: tdata_03 <- data_03 |>
      bind_cols(st_coordinates(data_03 |> st_transform(4326))) |>
      mutate(grupo = "Transporte",
             tipo = "Aeropuertos",
             info = "") |>
      select(grupo,
             tipo,
             nombre = BUSCA,
             CODMUN = CMUN,
             DIRECCION = DIRECCIÓN,
             X, Y,
             info) |> st_drop_geometry()
```

```
[68]: tdata_03 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	CODMUN <chr>	DIRECCION <chr>
A data.frame: 3 × 8	Transporte	Aeropuertos	Aeropuerto Adolfo Suárez-Madrid Barajas	079	Avda.
	Transporte	Aeropuertos	Aeropuerto Madrid-Cuatro Vientos	065	Ctra.
	Transporte	Aeropuertos	Aeropuerto Madrid-Torrejón de Ardoz	148	Ctra.

```
[69]: tdata_04 <- data_04 |>
      bind_cols(st_coordinates(data_04 |> st_transform(4326))) |>
```



```
mutate(grupo = "Transporte",
      tipo = "Estaciones de Cercanías",
      DIRECCION = CMUN,
      info = paste0("Líneas: ", LINEA)) |>
select(grupo,
      tipo,
      nombre = ETIQUETA,
      CODMUN = CMUN,
      DIRECCION,
      X, Y,
      info) |>
st_drop_geometry()
```

```
[70]: tdata_04 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	CODMUN <chr>	DIRECCION <chr>
A data.frame: 5 × 8	Transporte	Estaciones de Cercanías	Alcalá de Henares	005	005
	Transporte	Estaciones de Cercanías	Torrejón de Ardoz	148	148
	Transporte	Estaciones de Cercanías	San Fernando de Henares	049	049
	Transporte	Estaciones de Cercanías	Coslada	049	049
	Transporte	Estaciones de Cercanías	Vicálvaro	079	079

```
[71]: tdata_05 <- data_05 |>
      bind_cols(st_coordinates(data_05 |> st_transform(4326))) |>
      mutate(grupo = "Transporte",
            tipo = "Helisuperficies",
            DIRECCION = MUNICIPIO,
            info = paste0("Uso: ", USO, "; Gestor: ", GESTOR)) |>
      select(grupo,
            tipo,
            nombre = ETIQUETA,
            CODMUN = CMUN,
            DIRECCION,
            X, Y,
            info) |>
      st_drop_geometry()
```

```
[72]: tdata_05 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	CODMUN <chr>	DIRECCION <chr>
A data.frame: 5 × 8	Transporte	Helisuperficies	Helipuerto Hospital del Tajo	0133	A
	Transporte	Helisuperficies	Helipuerto Eventual de Fuentidueña del Tajo	0604	E
	Transporte	Helisuperficies	Helipuerto Eventual de Chinchón	0528	C
	Transporte	Helisuperficies	Campo de Ultraligeros de Valdelaguna	1570	V
	Transporte	Helisuperficies	Helipuerto Eventual de Estremera	0552	E

```
[73]: tdata_06 <- data_06 |>
      bind_cols(st_coordinates(data_06 |> st_transform(4326))) |>
      mutate(tipo = "Centros de Atención a Drogodependientes",
             grupo = "Salud",
             info = "") |>
      select(grupo,
             tipo,
             nombre = NOMBRE,
             CODMUN = CODMUN,
             DIRECCION,
             X, Y,
             info) |>
      st_drop_geometry()
```

```
[74]: tdata_06 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>
A data.frame: 5 × 8	Salud	Centros de Atención a Drogodependientes	Centro de Atención Integral a Drogod
	Salud	Centros de Atención a Drogodependientes	Centro de Atención Integral a Drogod
	Salud	Centros de Atención a Drogodependientes	Centro de Atención Integral a Drogod
	Salud	Centros de Atención a Drogodependientes	Centro de Atención Integral a Drogod
	Salud	Centros de Atención a Drogodependientes	Centro de Atención Integral a Drogod

```
[75]: tdata_07 <- data_07 |>
      bind_cols(st_coordinates(data_07 |> st_transform(4326))) |>
      mutate(tipo = "Centros de Especialidades",
             grupo = "Salud",
             info = paste0("CODZBS: ", CODZBS),
             DIRECCION = paste0(TIPOVIAL, NOMVIA, NUMERO)) |>
      select(grupo,
             tipo,
             nombre = ETIQUETA,
             CODMUN = CODMUN,
             DIRECCION,
             X, Y,
             info) |>
      st_drop_geometry()
```

```
[76]: tdata_07 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	CODMUN <chr>	D <chr>
A data.frame: 5 × 8	Salud	Centros de Especialidades	C.E. Moratalaz	0796	C
	Salud	Centros de Especialidades	C.E. Peña Prieta - Hermanos Sangr6	0796	A
	Salud	Centros de Especialidades	C.E. Vicente Soldevilla	0796	C
	Salud	Centros de Especialidades	C.E. Federica Montseny	0796	A
	Salud	Centros de Especialidades	C.E. Hermanos Garc6a Noblejas	0796	C

```
[77]: tdata_08 <- data_08 |>
      bind_cols(st_coordinates(data_08 |> st_transform(4326))) |>
      mutate(tipo = "Centros de Salud",
             grupo = "Salud",
             info = paste0("CODZBS: ", CODZBS),
             DIRECCION = paste0(TIPOVIAL, NOMVIA, NUMERO)) |>
      select(grupo,
             tipo,
             nombre = TIPOCEN2,
             CODMUN = CMUN,
             DIRECCION,
             X, Y,
             info) |>
      st_drop_geometry()
```

```
[78]: tdata_08 |> slice_head(n = 5)
```

A data.frame: 5 × 8

	grupo <chr>	tipo <chr>	nombre <chr>	CODMUN <chr>	DIRECCION <chr>
	Salud	Centros de Salud	Centro de Salud Carmen Calzado	005	CalleCarmen
	Salud	Centros de Salud	Centro de Salud Juan de Austria	005	AvenidaJuan
	Salud	Centros de Salud	Centro de Salud La Garena	005	CalleArturo S
	Salud	Centros de Salud	Centro de Salud Luis Vives	005	CalleLuis Viv
	Salud	Centros de Salud	Centro de Salud Manuel Merino	005	CalleManuel I

```
[79]: tdata_09 <- data_09 |>
      bind_cols(st_coordinates(data_09 |> st_transform(4326))) |>
      mutate(tipo = "Consultorios de Salud",
             grupo = "Salud",
             info = "",
             DIRECCION = paste0(TIPOVIA, NOMVIA, NUMERO)) |>
      select(grupo,
             tipo,
             nombre = ETIQUETA,
             CODMUN = CODMUNI,
             DIRECCION,
             X, Y,
             info) |>
      st_drop_geometry()
```

```
[80]: tdata_09 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	CODMUN <chr>	DIRECCION <chr>
A data.frame: 5 × 8	Salud	Consultorios de Salud	Consultorio de La Acebeda	0014	PlazaSan
	Salud	Consultorios de Salud	Consultorio de Ajalvir	0029	AvdaAgu
	Salud	Consultorios de Salud	Consultorio de Alameda del Valle	0035	PlazaSan
	Salud	Consultorios de Salud	Consultorio de El Álamo	0040	CalleFed
	Salud	Consultorios de Salud	Consultorio de Aldea del Fresno	0088	CalleSan

```
[81]: tdata_10 <- data_10 |>
  bind_cols(st_coordinates(data_10 |> st_transform(4326))) |>
  mutate(tipo = "Otros Centros de Salud",
         grupo = "Salud",
         info = "") |>
  select(grupo,
         tipo,
         nombre = NOMBRE,
         CODMUN,
         DIRECCION,
         X, Y,
         info) |>
  st_drop_geometry()
```

```
[82]: tdata_10 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	COORDENADAS <chr>
A data.frame: 5 × 8	Salud	Otros Centros de Salud	Instituto Provincial de Rehabilitación	079
	Salud	Otros Centros de Salud	Instituto Provincial de Oftalmología	079
	Salud	Otros Centros de Salud	Oficina Regional de Coordinación de Transplantes	079
	Salud	Otros Centros de Salud	Oficina Regional de Coordinación de Salud Mental	079
	Salud	Otros Centros de Salud	Oficina Regional de Coordinación Oncológica	079

```
[83]: tdata_11 <- data_11 |>
  bind_cols(st_coordinates(data_11 |> st_transform(4326))) |>
  mutate(tipo = "Centros de Salud Mental",
         grupo = "Salud",
         info = "",
         DIRECCION = paste0(TIPOVIAL, NOMVIA, NUMERO)) |>
  select(grupo,
         tipo,
         nombre = NOMBRE,
         CODMUN,
         DIRECCION,
         X, Y,
         info) |>
  st_drop_geometry()
```

```
[84]: tdata_11 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	CODMUN <chr>	DIRECCION <chr>
A data.frame: 5 × 8	Salud	Centros de Salud Mental	CSM Villa Vallecas	0796	CalleSan
	Salud	Centros de Salud Mental	CSM Coslada	0493	AvenidaL
	Salud	Centros de Salud Mental	CSM Salamanca	0796	CalleO'D
	Salud	Centros de Salud Mental	CSM Tetuán Infanto Juvenil	0796	CalleAgu
	Salud	Centros de Salud Mental	CSM Alcalá de Henarés I y II	0053	CalleOct

```
[85]: tdata_12 <- data_12 |>
      mutate(tipo = "Hospitales",
             grupo = "Salud",
             info = paste0("CAMAS: ", CAMAS, "; ", "Clase: ", `Clase de Centro`),
             DIRECCION = Dirección,
             CODMUN = as.character(`Cód. Municipio`)) |>
      select(grupo,
             tipo,
             nombre = `Nombre Centro`,
             CODMUN,
             DIRECCION,
             X, Y,
             info)
```

```
[86]: tdata_12 |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	CODMUN <chr>
A tibble: 5 × 8	Salud	Hospitales	HOSPITAL HM VALLES	280053
	Salud	Hospitales	HOSPITAL UNIVERSITARIO PRINCIPE DE ASTURIAS	280053
	Salud	Hospitales	HOSPITAL QUIRONSALUD SUR	280072
	Salud	Hospitales	HOSPITAL UNIVERSITARIO FUNDACION ALCORCON	280072
	Salud	Hospitales	HOSPITAL DEL TAJO S.A.	280133

```
[87]: data <-
      bind_rows(tdata_01, tdata_02, tdata_03,
               tdata_04, tdata_05, tdata_06,
               tdata_07, tdata_08, tdata_09,
               tdata_10, tdata_11, tdata_12)
```

```
[88]: glimpse(data)
```

```
Rows: 1,633
Columns: 8
$ grupo    <chr> "Transporte", "Transporte", "Transporte",
"Transporte", "Tra...
$ tipo     <chr> "Intercambiadores", "Intercambiadores",
"Intercambiadores", ...
$ nombre   <chr> "Grandes Intercambiadores Plaza Elíptica",
"Grandes Intercam...
$ CODMUN   <chr> "079", "079", "079", "079", "079", "079",
```

```
"079", "079", "079..."
$ DIRECCION <chr> "Plaza Elíptica s/n", "Calle Princesa, 89",
"Estación Prínci..."
$ X <dbl> -3.716577, -3.719508, -3.719560, -3.689044,
-3.676731, -3.68...
$ Y <dbl> 40.38540, 40.43474, 40.42080, 40.46719,
40.43797, 40.47198, ...
$ info <chr> "", "", "", "", "", "", "", "", "", "", "", "",
"", "", "", "", ...
```

```
[89]: data |> slice_head(n = 5)
```

	grupo <chr>	tipo <chr>	nombre <chr>	CODMU <chr>
A data.frame: 5 × 8	Transporte	Intercambiadores	Grandes Intercambiadores Plaza Elíptica	079
	Transporte	Intercambiadores	Grandes Intercambiadores Moncloa	079
	Transporte	Intercambiadores	Grandes Intercambiadores Príncipe Pío	079
	Transporte	Intercambiadores	Grandes Intercambiadores Plaza de Castilla	079
	Transporte	Intercambiadores	Grandes Intercambiadores Avenida de América	079

## 0.4 Synthetic Data Generation

Estos datos no requieren tareas de este tipo.

## 0.5 Fake Data Generation

Estos datos no requieren tareas de este tipo.

## 0.6 Open Data

Los archivos proceden de fuentes abiertas en tareas anteriores.

## 0.7 Data Save

Este proceso, puede copiarse y repetirse en aquellas partes del notebbok que necesiten guardar datos. Recuerde cambiar las cadenas añadida del fichero para diferenciarlas

Identificamos los datos a guardar

```
[90]: data_to_save <- data
```

Estructura de nombre de archivos:

- Código del caso de uso, por ejemplo "CU\_04"
- Número del proceso que lo genera, por ejemplo "\_05".
- Número de la tarea que lo genera, por ejemplo "\_01"
- En caso de generarse varios ficheros en la misma tarea, llevarán \_01 \_02 ... después
- Nombre: identificativo de "properData", por ejemplo "\_zonasgeo"
- Extensión del archivo

Ejemplo: "CU\_04\_05\_01\_01\_zonasgeo.json, primer fichero que se genera en la tarea 01 del proceso 05 (Data Collection) para el caso de uso 04 (vacunas)

Importante mantener los guiones bajos antes de proceso, tarea, archivo y nombre

### 0.7.1 Proceso 05

```
[91]: caso <- "CU_18"  
      proceso <- '_05'  
      tarea <- "_02"  
      archivo <- ""  
      proper <- "_infraestructuras"  
      extension <- ".csv"
```

OPCION A: Uso del paquete “tcltk” para mayor comodidad

- Buscar carpeta, escribir nombre de archivo SIN extensión (se especifica en el código)
- Especificar sufijo2 si es necesario
- Cambiar datos por datos\_xx si es necesario

```
[92]: # file_save <- paste0(caso, proceso, tarea, tcltk::tkgetSaveFile(), proper,   
      ↪extension)  
      # path_out <- paste0(oPath, file_save)  
      # write_csv(data, path_out)  
  
      # cat('File saved as: ')  
      # path_out
```

OPCION B: Especificar el nombre de archivo

- Los ficheros de salida del proceso van siempre a Data/Output/.

```
[93]: file_save <- paste0(caso, proceso, tarea, archivo, proper, extension)  
      path_out <- paste0(oPath, file_save)  
      write_csv(data, path_out)  
  
      cat('File saved as: ')  
      path_out
```

File saved as:

'Data/Output/CU\_18\_05\_02\_infraestructuras.csv'

**Copia del fichero a Input** Si el archivo se va a usar en otros notebooks, copiar a la carpeta Input

```
[94]: path_in <- paste0(iPath, file_save)  
      file.copy(path_out, path_in, overwrite = TRUE)
```

TRUE

## 0.8 Main Conclusions

List and describe the general conclusions of the analysis carried out.

### 0.8.1 Prerequisites

This working code needs the following conditions:

- For using the interactive selection of file, the {tcltk} package must be installed. It is not needed in production.
- The {dplyr}, {readr} and {sf} packages must be installed.
- The data paths `Data/Input` and `Data/Output` must exist (relative to the notebook path)

### 0.8.2 Configuration Management

This notebook has been tested with the following versions of R and packages. It cannot be assured that later versions work in the same way: \* R 4.2.2 \* tcltk 4.2.2 \* dplyr 1.0.10 \* readr 2.1.3

### 0.8.3 Data structures

#### Objeto data

- Los datos de origen contienen información diversa que se ha homogeneizado en el archivo de salida
- Hay 1633 filas con las variables:
  - grupo
  - tipo
  - nombre
  - CODMUN
  - DIRECCION
  - X
  - Y
  - info

#### Observaciones generales sobre los datos

- Las infraestructuras de metro están por “boca de metro”. Al aplicar los modelos, ver si conviene agrupar por estación.
- La proyección de origen de la georeferenciación es ETRS89 (CRS 4258). En todo el proyecto se está utilizando prioritariamente WGS 84 (CRS 4326), por lo que se hace esta transformación. Si en algún momento esto cambia habría que modificar lo que corresponda en el apartado de transformación de datos

### 0.8.4 Consideraciones para despliegue en piloto

- Los datos de origen fueron obtenidos de <https://gestiona.comunidad.madrid/nomecalles/DescargaBDTCorte>. en fichero zip con formato .shp y descomprimido en la carpeta `Data/Input`
- Si la información geográfica cambia, se debería actualizar el fichero y volver a ejecutar todos los procesos.

### 0.8.5 Consideraciones para despliegue en producción

- Se deben crear los procesos ETL en producción necesarios para que los datos de entrada estén actualizados



## 0.9 Main Actions

**Acciones done** Indicate the actions that have been carried out in this process

- Se han transformado las geometrías a WGS 84 (CRS 4326)

**Actions to perform** Indicate the actions that must be carried out in subsequent processes

- Asignar sección censal
- Asignar capacidad de cada infraestructura
- Obtener datos de ocupación por día de las infraestructuras
- Agregar por sección censal

## 0.10 CODE TO DEPLOY (PILOT)

A continuación se incluirá el código que deba ser llevado a despliegue para producción, dado que se entiende efectúa operaciones necesarias sobre los datos en la ejecución del prototipo

Description

- No hay nada que desplegar en el piloto, ya que estos datos son estáticos o en todo caso cambian con muy poca frecuencia, altamente improbable durante el proyecto.

CODE

[95]: `# incluir código`