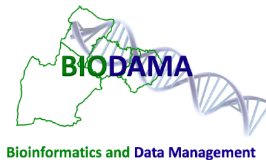


GEO-CIBER

Manejo de datos generados en R





Pablo Fernández Navarro

e-mail: biodama.cne@isci.iii.es

Unidad de Epidemiología del Cáncer y Ambiental, Centro Nacional de
Epidemiología, Instituto de Salud Carlos III

Material

- Presentaciones en pdf
- Datos
- App para windows/mac/linux

Descarga de material:



https://github.com/biodama/GEO_CIBER

Estructura del taller

1. Paquetes necesarios
2. Importación de direcciones geocodificadas con GEO-CIBER
3. Visualizamos los datos generales
4. Importación de direcciones geocodificadas con GEO-CIBER
5. Visualizamos los datos adicionales
6. Control de calidad básico Scores
7. Control de calidad básico Portales
8. Conversión a datos espaciales
9. Visualización datos espaciales
10. Creación de kmls
11. Superposición de puntos en mapa
12. Unión datos epidemiológicos
13. Datos población en mapa
14. Creación datos en rejilla
15. Exportación de datos

1. Instalación de paquetes de R necesarios para el Taller

```
install.packages("scales")
```

```
install.packages("sf")
```

```
install.packages("tmaptools")
```

```
install.packages("tmap")
```

```
install.packages("ggplot2")
```

```
install.packages("sp")
```

```
install.packages("openxlsx")
```

```
install.packages("rgdal")
```

2.Importación de direcciones geocodificadas con GEO-CIBER

Datos generales

```
rm(list=ls())
gc()

##           used (Mb) gc trigger   (Mb) limit (Mb) max used (Mb)
## Ncells 1322516 70.7   2744952 146.6      NA   1468538 78.5
## Vcells 2259053 17.3   8388608  64.0     32768  3055653 23.4

require("openxlsx")

datos<-read.xlsx("res/2023-12-07_resultado_geo_ciber.xlsx")

dim(datos)

## [1] 112  12
```

3. Visualizamos los datos generales

```
head(datos[,c("ID", "lat_geo_ciber",  
              "long_geo_ciber", "geo_ciber_score")])
```

| ## | | ID | lat_geo_ciber | long_geo_ciber | geo_ciber_score |
|------|--|--------------|------------------|-------------------|-----------------|
| ## 1 | | 101310022153 | 39.8925719951015 | -5.54388882480936 | 3 |
| ## 2 | | 101310126019 | 39.8918143315662 | -5.53686783828005 | 3 |
| ## 3 | | 101310242613 | 39.8909887215523 | -5.54192667587535 | 3 |
| ## 4 | | 101310001579 | 39.8930986398904 | -5.54102624676381 | 3 |
| ## 5 | | 101310001839 | 39.8894674814392 | -5.5411978212224 | 3 |
| ## 6 | | 101310003991 | 39.895377002147 | -5.53914153395459 | 3 |

4.Importación de direcciones geocodificadas con GEO-CIBER

Datos adicionales

```
datos_ad<-read.xlsx("res/2023-12-07_resultado_geo_ciber_adicional.xlsx")
```

```
dim(datos_ad)
```

```
## [1] 112 32
```


5. Visualizamos los datos adicionales

```
head(datos_ad[,c("ID", "portalNumber_carto", "province_carto")])
```

| ## | ID | portalNumber_carto | province_carto |
|------|--------------|--------------------|----------------|
| ## 1 | 101310022153 | 16 | Cáceres |
| ## 2 | 101310126019 | 27 | Cáceres |
| ## 3 | 101310242613 | 17 | Cáceres |
| ## 4 | 101310001579 | 8 | Cáceres |
| ## 5 | 101310001839 | 6 | Cáceres |
| ## 6 | 101310003991 | 6 | Cáceres |

6.Control de calidad básico Scores (1/3)

```
# Check scores
```

```
table(datos$"geo_ciber_score")
```

```
##
```

```
##      2      3
```

```
##      1 111
```

```
table(datos_ad$"comb_score",datos_ad$"combinaciones")
```

```
##
```

```
##              2      3
```

```
## 0_1          1      0
```

```
## 2_0          0     11
```

```
## 2_1          0    100
```

6.Control de calidad básico Scores (2/3)

```
check_score<-which(datos_ad$"combinaciones"==2)
```

```
datos_ad[check_score,c("ID","NUM","portalNumber_carto",  
                        "comb_score","combinaciones")]
```

```
##           ID NUM portalNumber_carto comb_score combinaciones  
## 69 101310251555    9                <NA>         0_1          2
```

```
datos_ad[check_score,c("TIPO_VIA","DIRECCION","MUNICIPIO",  
                        "PROVINCIA","province_carto")]
```

```
##           TIPO_VIA DIRECCION                MUNICIPIO PROVINCIA province_carto  
## 69 DESCONOCIDO      -997 NAVALMORAL DE LA MATA    CÁCERES          <NA>
```

6.Control de calidad básico Scores (3/3)

```
# Eliminamos el registro
```

```
delete_reg<-which(datos_ad$"combinaciones"==2)
```

```
datos<-datos[-delete_reg,]
```

```
datos_ad<-datos_ad[-delete_reg,]
```

```
row.names(datos)<-NULL
```

```
row.names(datos_ad)<-NULL
```

```
dim(datos)
```

```
## [1] 111 12
```

```
dim(datos_ad)
```

```
## [1] 111 32
```

7.Control de calidad básico Portales (1/2)

```
missing_num<-which(is.na(datos_ad$"portalNumber_carto"))
datos_ad$"portalNumber_carto"[missing_num]<-99999

check_num<-which(datos_ad$"NUM"!=datos_ad$"portalNumber_carto")

datos_ad[check_num,c("ID","NUM","portalNumber_carto",
                     "comb_score","combinaciones")]
```

| ## | ID | NUM | portalNumber_carto | comb_score | combinaciones |
|-------|--------------|-----|--------------------|------------|---------------|
| ## 68 | 101310173874 | 16 | 14 | 2_0 | 3 |
| ## 77 | 101310188828 | 8 | 99999 | 2_1 | 3 |

```
datos_ad[check_num,c("TIPO_VIA","DIRECCION","MUNICIPIO",
                     "PROVINCIA","province_carto")]
```

| ## | TIPO_VIA | DIRECCION | MUNICIPIO | PROVINCIA | province_carto |
|-------|----------|-------------|-----------------------|-----------|----------------|
| ## 68 | CALLEJON | FUENTENUEVA | NAVALMORAL DE LA MATA | CÁCERES | Cáceres |
| ## 77 | CALLEJON | FUENTENUEVA | NAVALMORAL DE LA MATA | CÁCERES | <NA> |

7.Control de calidad básico Portales(2/2)

IR AL ARCHIVO KML GENERAO POR GEO-CIBER

8. Conversion a datos espaciales (1/2)

```
require(sf)
require(rgdal)
```

```
EPSG <- make_EPSG()
EPSG[EPSG$"note"%in%"WGS 84",]
```

```
##          code   note                                     prj4
## 2180 4326 WGS 84                                     +proj=longlat +datum=WGS84 +no_defs +type=crs
## 2756 4978 WGS 84 +proj=geocent +datum=WGS84 +units=m +no_defs +type=crs
## 2757 4979 WGS 84                                     +proj=longlat +datum=WGS84 +no_defs +type=crs
##          prj_method
## 2180      (null)
## 2756      (null)
## 2757      (null)
```

```
# st_as_sf
```

```
datos_sf <- sf::st_as_sf(datos,
                          coords = c("long_geo_ciber", "lat_geo_ciber"),
                          crs = 4326)
```

8. Conversion a datos espaciales (2/2)

```
# st_read y st_transform
```

```
navalmoral.shp <- st_read("datos/mapa_naval.shp")
```

```
## Reading layer `mapa_naval' from data source
```

```
##   `/Users/pfernandezn/Desktop/geo_ciber/datos/mapa_naval.shp'
```

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

```
## Simple feature collection with 12 features and 16 fields
```

```
## Geometry type: POLYGON
```

```
## Dimension:      XY
```

```
## Bounding box:   xmin: 271677.7 ymin: 4415382 xmax: 288816.6 ymax: 4430892
```

```
## Projected CRS: ETRS89 / UTM zone 30N
```

```
navalmoral.shp<-st_transform(navalmoral.shp, 4326)
```


9. Visualización datos espaciales

```
# st_geometry
```

```
plot(st_geometry(navalmoral.shp))
```

```
plot(st_geometry(datos_sf), pch=3, col="red", add=T)
```



10.Creación de kmls

```
# st_write
```

```
st_write(datos_sf, "res/datos_geocodificados.kml",  
         driver = "kml", delete_dsn = TRUE)
```

```
## Deleting source `res/datos_geocodificados.kml' using driver `kml'  
## Writing layer `datos_geocodificados' to data source  
## `res/datos_geocodificados.kml' using driver `kml'  
## Writing 111 features with 10 fields and geometry type Point.
```

```
st_write(navalmoral.shp, "res/mapa_navalmoral.kml",  
         driver = "kml", delete_dsn = TRUE)
```

```
## Deleting source `res/mapa_navalmoral.kml' using driver `kml'  
## Writing layer `mapa_navalmoral' to data source  
## `res/mapa_navalmoral.kml' using driver `kml'  
## Writing 12 features with 16 fields and geometry type Polygon.
```

11.Superposición de puntos en mapa

```
# st_intersects  
  
indice<-unlist(st_intersects(datos_sf, navalmoral.shp))  
  
datos_sf$CUSEC<-navalmoral.shp$CUSEC[indice]
```

12.Unión datos epidemiológicos

```
# merge
```

```
datos.epi<-read.xlsx("datos/datos.epi.xlsx")
```

```
head(datos.epi)
```

```
##           ID    sexo edad  edad.gr
## 1 101310022153 Hombre   40  (39,44]
## 2 101310126019  Mujer   86  (84,100]
## 3 101310242613  Mujer   47  (44,49]
## 4 101310001579  Mujer   23  (19,24]
## 5 101310001839  Mujer   39  (34,39]
## 6 101310003991  Mujer   47  (44,49]
```

```
datos_sf<-merge(datos_sf,datos.epi,by="ID")
```

13.Datos población en mapa

```
# merge

pob<-read.xlsx("datos/poblacion_cusec_caceres.xlsx")
pob$ambos<-pob$Hombre+pob$Mujer

navalmoral.shp<-merge(navalmoral.shp,pob,by="CUSEC")
```

14.Creación datos en rejilla (1/2)

```
# aggregate
```

```
casos.sec<-aggregate(ID~CUSEC, data=datos_sf, FUN =length)
```

```
casos.sec
```

```
##           CUSEC ID
## 1  1013101001 18
## 2  1013101002 12
## 3  1013101003 15
## 4  1013101004  8
## 5  1013102001  4
## 6  1013102002 12
## 7  1013102003 14
## 8  1013102004  9
## 9  1013102005 12
## 10 1013102006  5
## 11 1013102007  2
```

14.Creación datos en rejilla (2/2)

```
# merge
```

```
casos.shp <- merge(naualmoral.shp,casos.sec,all.x=T,all.y=T,by="CUSEC")
```

```
casos.shp$TC<-casos.shp$ID/casos.shp$ambos
```

15.Exportación de datos

```
# save  
  
casos<-datos_sf  
  
save(navalmoral.shp,casos,file="datos/datos_procesados.RData")
```