

## Guía laboratorio 1: Introducción a los SIG

### Modulo 1: Manejo de shapefiles

#### Descarga de datos

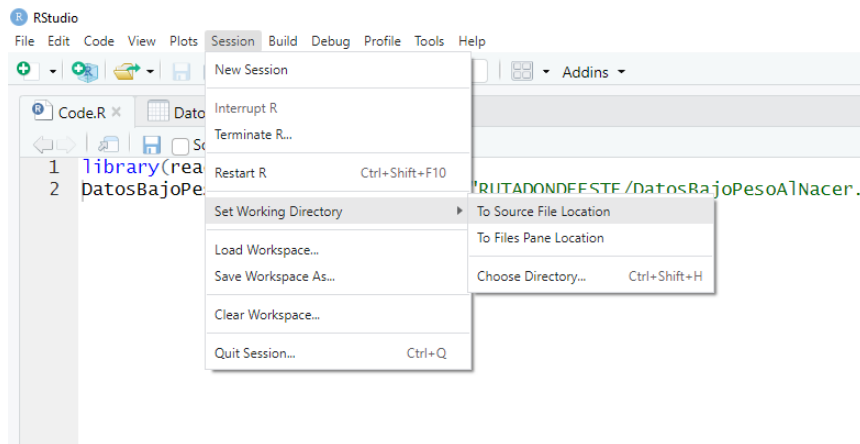
1. Descargue de el SIVIGILA (<https://portalsivigila.ins.gov.co/Paginas/Vigilancia-Rutinaria.aspx>) los datos de vigilancia rutinaria para el 2022. Ojo que donde dice descargar 995 (Morbilidad por IRA) y 998 (Morbilidad por EDA). Baje el Excel y en este depure los datos a nivel departamental que le quede una base de datos que diga departamento, código de departamento y total de casos para el 2022 para bajo peso al nacer.

DEPARTAMENTO	ID DANE	ENFERMEDAD	CASOS
AMAZONAS	91	BAJO PESO AL NACER	25
ANTIOQUIA	05	BAJO PESO AL NACER	1152
ARAUCA	81	BAJO PESO AL NACER	91
ATLANTICO	08	BAJO PESO AL NACER	321
BOGOTA	11	BAJO PESO AL NACER	52
BOLIVAR	13	BAJO PESO AL NACER	438
BOYACA	15	BAJO PESO AL NACER	379
CALDAS	17	BAJO PESO AL NACER	198
CAQUETA	18	BAJO PESO AL NACER	74
CASANARE	85	BAJO PESO AL NACER	111
CAUCA	19	BAJO PESO AL NACER	394
CESAR	20	BAJO PESO AL NACER	330
CHOCO	27	BAJO PESO AL NACER	121
CORDOBA	23	BAJO PESO AL NACER	482
CUNDINAMARCA	25	BAJO PESO AL NACER	735
EXTERIOR	01	BAJO PESO AL NACER	54
GUAINIA	94	BAJO PESO AL NACER	15
GUAJIRA	44	BAJO PESO AL NACER	276
GUAVIARE	95	BAJO PESO AL NACER	24
HUILA	41	BAJO PESO AL NACER	324
MAGDALENA	47	BAJO PESO AL NACER	310
META	50	BAJO PESO AL NACER	177
NARIÑO	52	BAJO PESO AL NACER	348
NORTE SANTANDER	54	BAJO PESO AL NACER	287
PUTUMAYO	86	BAJO PESO AL NACER	89
QUINDIO	63	BAJO PESO AL NACER	96
RISARALDA	66	BAJO PESO AL NACER	174
SAN ANDRES	88	BAJO PESO AL NACER	20
SANTANDER	68	BAJO PESO AL NACER	470
SUCRE	70	BAJO PESO AL NACER	302
TOLIMA	73	BAJO PESO AL NACER	243
VALLE	76	BAJO PESO AL NACER	440

2. Descargue el marco geoestadístico nacional (<https://www.dane.gov.co/files/geoportal-provisional/index.html>) para el 2021. En este archivo nos vamos a enfocar para esta práctica en el archivo de departamentos que esta en ADMINISTRATIVO. Tenemos que usar todos los archivos que acompañan al .shp que tiene los departamentos del país.
3. Descargue de la página del DANE (<https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion>) las proyecciones de población **por área** para el 2022 con base en el censo nacional de población y vivienda (CNPV) 2018. Copie los datos de interés a una tabla nueva.

#### Carga de datos en R

1. Primero asegúrese de crear una carpeta donde tenga todos los datos, y haga esta carpeta como su *working directory*, el cual será la carpeta base donde R buscare todos los archivos.

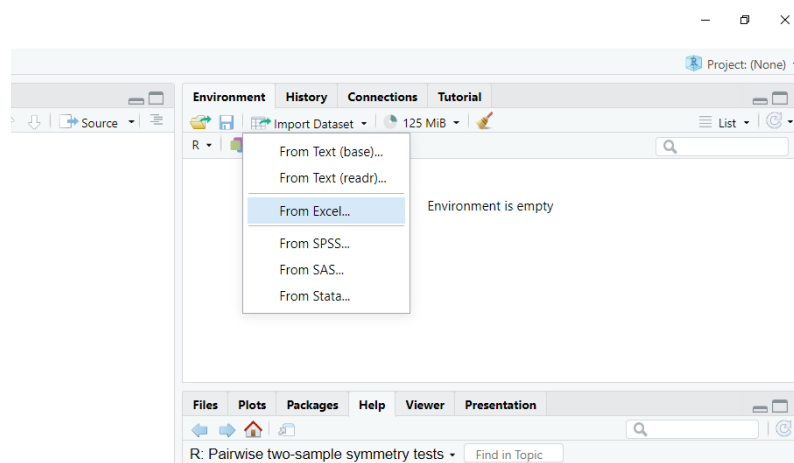


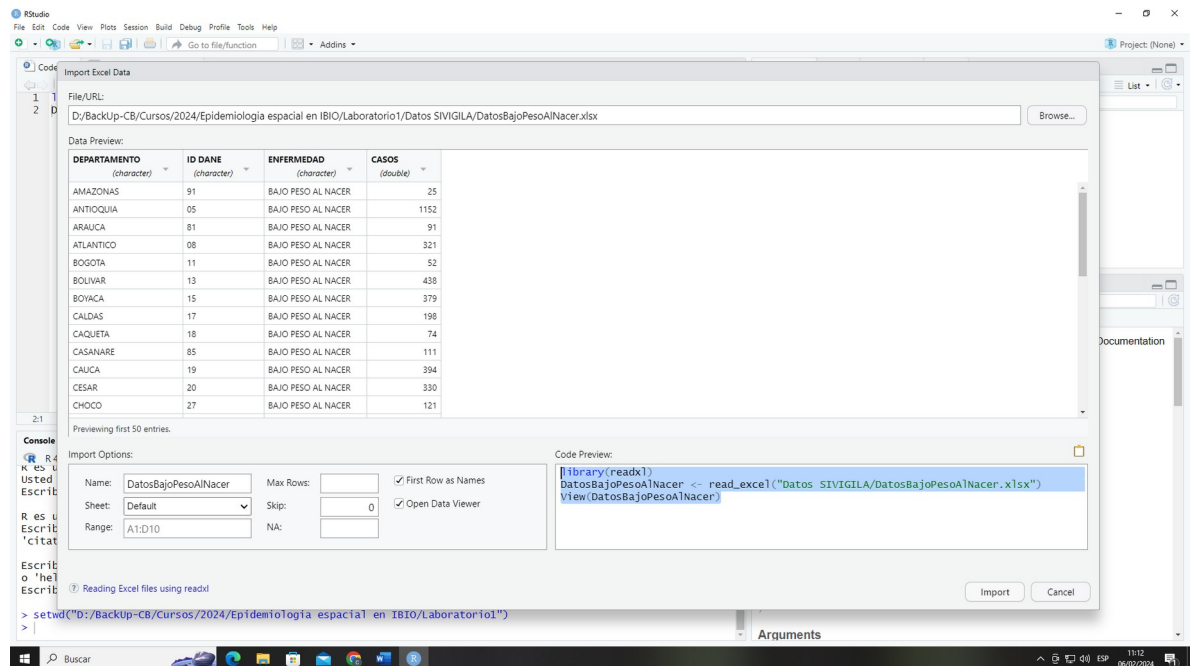
Acá *To Source File Location* pone el working directory a la carpeta donde tiene guardado el código que tiene abierto. *To Files Pane Location* lo pone en el navegador que encuentra en la ventana files de la esquina inferior izquierda, y *Choose Directory* le permite navegar y elegir este directorio. Copie el código que le sale a su código para tener el comando que define su working directory y así siempre pueda seleccionarlo sin necesidad de volver a hacer este procedimiento.

`setwd("D:/BackUp-CB/Cursos/2024/Epidemiologia espacial en IBIO/Laboratorio1")`

En este comando tiene que insertar como un *string* la dirección de la carpeta que será su working directory. Por esto entra entre dos comillas dobles.

2. Cargue las bases de datos de Excel usando el módulo de *ImportDataset* que viene por default en R. Recuerde cargarlo usando la librería *readxl*, y copie el código para poderlo cargar inmediatamente las veces posteriores sin tener que hacer este procedimiento.





De esta manera, su código hasta esta parte se debería ver de la siguiente forma:

```
library(readxl) #CARGAMOS LIBRERIAS
```

```
setwd("D:/BackUp-CB/Cursos/2024/Epidemiologia espacial en  
IBIO/Laboratorio1")#ELEGIMOS WORKING DIRECTORY
```

```
DatosBajoPesoAlNacer <- read_excel("Datos  
SIVIGILA/DatosBajoPesoAlNacer.xlsx")#CARGAMOS DATOS DE BAJO PESO AL  
NACER
```

```
DatosPoblación <-  
read_excel("DatosPoblacion/DatosPoblacion.xlsx")#CARGAMOS DATOS DE  
población
```

Recuerde que acá el # crea comentarios, por lo cual estas partes del código no se toman en cuenta como código por el computador. En mi caso, yo tengo una carpeta llamada Laboratorio1 que es donde tengo el código, y tengo en esa carpeta dos carpetas: Datos SIVIGILA y DatosPoblacion, donde tengo los datos que descargamos previamente de Excel. Por esto cuando cargo los datos con la función `read_excel`, pongo esto en el texto *nombre de la carpeta/nombre del archivo.xlsx*.

Por otro lado, en la esquina superior derecha de rstudio está en *environment*, donde muestra todas las variables. Tras correr estos comandos podemos ver que creamos dos variables para cada set de datos que pueden ser observadas en este panel:

Environment	History	Connections	Tutorial
Import Dataset	141 MiB		
R	Global Environment		
Data			
DatosBajoPesoAlN...	34 obs. of 4 variables		
DatosPoblación	2079 obs. of 5 variables		

Ahora, vamos a observar los datos con el comando `View` para asegurarnos que hayan compilado bien:

`View(DatosBajoPesoAlNacer)` #Vemos datos bajo peso

`View(DatosPoblación)` #Vemos datos población

	DEPARTAMENTO	ID DANE	ENFERMEDAD	CASOS
1	AMAZONAS	91	BAJO PESO AL NACER	25
2	ANTIOQUIA	05	BAJO PESO AL NACER	1152
3	ARAUCA	81	BAJO PESO AL NACER	91
4	ATLANTICO	08	BAJO PESO AL NACER	321
5	BOGOTA	11	BAJO PESO AL NACER	52
6	BOLIVAR	13	BAJO PESO AL NACER	438
7	BOYACA	15	BAJO PESO AL NACER	379
8	CALDAS	17	BAJO PESO AL NACER	198
9	CAQUETA	18	BAJO PESO AL NACER	74
10	CASANARE	85	BAJO PESO AL NACER	111
11	CAUCA	19	BAJO PESO AL NACER	394
12	CESAR	20	BAJO PESO AL NACER	330
13	CHOCO	27	BAJO PESO AL NACER	121
14	CORDOBA	23	BAJO PESO AL NACER	482
15	CUNDINAMARCA	25	BAJO PESO AL NACER	735

	DP	DPNOM	AÑO	ÁREA GEOGRÁFICA	Población
1	05	Antioquia	2020	Cabecera	5250157
2	05	Antioquia	2020	Centros Poblados y Rural Disperso	1399244
3	05	Antioquia	2020	Total	6649401
4	05	Antioquia	2021	Cabecera	5350556
5	05	Antioquia	2021	Centros Poblados y Rural Disperso	1375663
6	05	Antioquia	2021	Total	6726219
7	05	Antioquia	2022	Cabecera	5424921
8	05	Antioquia	2022	Centros Poblados y Rural Disperso	1362925
9	05	Antioquia	2022	Total	6787846
10	05	Antioquia	2023	Cabecera	5494521
11	05	Antioquia	2023	Centros Poblados y Rural Disperso	1353839
12	05	Antioquia	2023	Total	6848360
13	05	Antioquia	2024	Cabecera	5551824
14	05	Antioquia	2024	Centros Poblados y Rural Disperso	1351897
15	05	Antioquia	2024	Total	6903721
16	05	Antioquia	2025	Cabecera	5600952

Ahora vamos a cargar los datos del marco geoestadístico nacional. Para esto usamos el comando `vect()` de la librería *terra*.

```
install.packages("terra") #instalamos paquete, recuerden comentar #esta
línea después ya que solo es necesario instalar una ves.
```

```
library(terra) #compilamos el paquete
```

```
Map=vect("MapaDeptoCOL/MGN_DPTO_POLITICO.shp")#Cargamos el mapa
```

Recordemos que usamos el comando `vect()` ya que el mapa es un shapefile, ósea es un mapa vectorial. Vamos a hacer una gráfica del mapa.

```
plot(Map) #Gráficamos el mapa
```

Ahora, vamos a ver el Coordinate Reference System del mapa (CRS).

```
crs(Map) #Vemos el crs del mapa
```

```
"GEOGCRS["MAGNA-SIRGAS",\n  DATUM["Marco Geocentrico Nacional de
Referencia",\n    ELLIPSOID["GRS 1980",6378137,298.257222101,\n  LENGTHUNIT["metre",1]],\n  PRIMEM["Greenwich",0,\n  ANGLEUNIT["degree",0.0174532925199433]],\n  CS["ellipsoidal",2],\n  AXIS["geodetic latitude (Lat)",north,\n  ORDER[1],\n  ANGLEUNIT["degree",0.0174532925199433]],\n  AXIS["geodetic
longitude (Lon)",east,\n  ORDER[2],\n  ANGLEUNIT["degree",0.0174532925199433]],\n  USAGE[\n  SCOPE["Horizontal component of 3D system.",\n  AREA["Colombia -
onshore and offshore. Includes San Andres y Providencia, Malpelo Islands,
Roncador Bank, Serrana Bank and Serranilla Bank.",\n  BBOX[-4.23,-
84.77,15.51,-66.87]],\n  ID["EPSG",4686]]"
```

Vemos que el resultado (Ver texto en gris) tiene todas las unidades del CRS, junto al DATUM que es MAGNA-SIRGAS y nuestro EPSG es el 4686. Investigue un poco sobre este CRS. Por esto, si usted carga otro mapa internacional en otro CRS debería proyectar. Si queremos ver el dataframe asociado al mapa:

```
View(as.data.frame(Map)) #Vemos el crs del mapa
```

Ahora, vamos a integrar los datos reportados por el SIVIGILA. Para esto, tenemos que hacer un cruce de información usando el ID asignado por el DANE. Para esto, podemos usar el comando `%in%` que nos retorna un vector de verdaderos y falsos en las posiciones que comparte el argumento previo con el posterior al comando.

```
Map$DPTO_CCDGO %in% DatosBajoPesoAINacer$`ID DANE` #Vemos el
vect#or de verdaderos y falsos según los valores que compartan.
```

Como arroja todo `true` quiere decir que todos los códigos de departamento que están en nuestro mapa se encuentran en la base de datos del SIVIGILA, por lo cual podemos realizar la asignación. Podríamos incluso usar este vector para hacer un subset de datos y detectar cuales códigos están y cuales no están:

```
Map$DPTO_CCDGO[Map$DPTO_CCDGO %in% DatosBajoPesoAINacer$`ID
DANE`] #Aca usamos el vector de verdaderos y falsos para hacer un subs#set
y saber que códigos del dane de nuestro DANE están en la base del SIVIGILA.
```

Si queremos ver cuales no están, usamos el comando `!` que niega el verdadero o el falso sobre el vector de verdaderos y falsos que nos retorna el `%in%`. De esta manera, obtenemos los códigos del dane del mapa que no están en los datos del SIVIGILA.

```
Map$DPTO_CCDGO[!Map$DPTO_CCDGO %in% DatosBajoPesoAINacer$`ID
DANE`] #Aca usamos el vector de verdaderos y falsos para hacer un subs#set
y saber que códigos del dane de nuestro DANE NO están en la base del
SIVIGILA.
```

Vemos que retorna `character(0)` ya que todos los códigos del dane del mapa están en la base de datos del SIVIGILA. Ahora, podemos hacer un recorrido para pasar los datos de manera manual. Ojo que luego veremos métodos para usar algo más sencillo.

```
Map$CasosBajoPesoAINacer=NA #Inicializamos la columna

#Hacemos el recorrido. El contador i va sobre los valores de la c#olumna del
código del dane del mapa.
for (i in Map$DPTO_CCDGO){

  Map$CasosBajoPesoAINacer[Map$DPTO_CCDGO==i]=DatosBajoPesoAINacer$
CASOS[DatosBajoPesoAINacer$`ID DANE`==i]
  #Aca los corchetes nos permiten hacer el subset.
}
```

Ahora vamos a graficar el mapa coloreado según los casos con el comando `plot`.

```
plot(Map,"CasosBajoPesoAINacer",plg=list(title="CasosBajoPesoAINacer"))#H
acemos el plot
```

En este `plot`, entre comillas va el nombre de la columna del mapa sobre la cual queremos colorear, y en `plg` va una lista de parámetros donde únicamente usamos `title` que es el título de la leyenda.

## Laboratorio 1

1. ¿Del mapa que muestra los casos, que región la más afectado? ¿Concluir que estas regiones son las de mayor *riesgo* tendría algún tipo de sesgo?
2. Ahora pase los datos de población para el año 2022 a su mapa. Recuerde que debe hacer un subset de su base de datos de población primero únicamente el año 2022 para todas las áreas (Rural, Urbana y Total), luego revisar que para todos sus departamentos del mapa estén los datos y ahí si podrá hacer el cruce de datos con `for`.
3. Calcule la incidencia por cada 100.000 habitantes para el 2022 y justifique que área de población entre rural urbana y total va a usar para calcular este índice, haga un mapa y contraste los resultados con los obtenidos únicamente mirando los casos.