

Introducción a los datos espaciales

Jorge Juvenal Campos Ferreira



Programa interdisciplinario
de Ciencia de Datos

Junio 2024



Sobre mí:



M.C. JORGE JUVENAL CAMPOS FERREIRA.

Analista de datos en **México ¿Cómo vamos?** Y
en **Fundación Novagob México**

Educación Formal:



Licenciatura:

Ingeniería en Irrigación por la Universidad Autónoma Chapingo. (2009–2014).



Maestría:

Maestría en Economía por El Colegio de México (2016–2018).

Contacto:



GitHub: JuveCampos



LinkedIn: Jorge Juvenal Campos Ferreira



Twitter: @JuvenalCamposF



IG: juvenalcampos.dataviz

Objetivos de esta parte del curso:



En esta parte del curso veremos los siguientes temas:

Tema 01. Introducción a los datos espaciales.

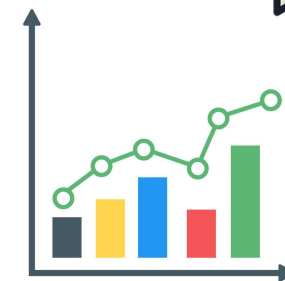
Tema 02. Visualización de datos espaciales

Tema 03. Manejo de datos espaciales.

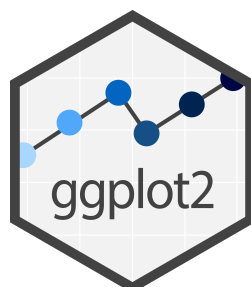
Tema 04. Procesos con datos espaciales

Tema 05. Visualización interactiva

Tema 06. Introducción a las aplicaciones en r shiny



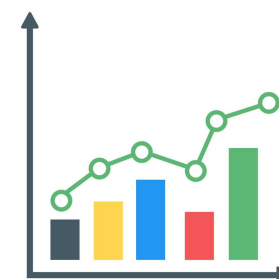
+



+



+



=



Objetivos de la sesión:

1. Que los estudiantes conozcan las **principales herramientas** de manejo de información geográfica
2. Que los estudiantes conozcan **los principales formatos** en los que se almacena la información geográfica y como manejar cada uno de ellos.
3. Que los estudiantes conozcan los siguientes conceptos:
 - a. Sistema de coordenadas de referencia
 - c. Información vectorial
 - d. Información raster
4. Que los estudiantes **conozcan las principales librerías** de R para manejar datos geográficos: *{sf}*, *{sp}*, *{cartogram}*, *{leaflet}*, *{mxmaps}*, *etc.*
5. Que los estudiantes **aprendan a cargar información geográfica** en sus sesiones de R
6. Que los estudiantes **aprendan a explorar los datos espaciales** en sus sesiones de R.
7. Que los estudiantes **aprendan a exportar sus datos espaciales** a archivos externos.
8. Que los estudiantes **aprendan a identificar el nivel de agregación geográfico** de sus bases de datos
9. Que los estudiantes **aprendan de donde pueden sacar información geográfica** para trabajar en R.



**Herramientas de manejo de
información geográfica.**

Herramientas de manejo de información geográfica.



ArcGIS: Desarrollado por Esri, es una de las herramientas SIG más utilizadas en el mundo. Ofrece una amplia gama de funcionalidades para el análisis espacial, la visualización y la gestión de datos geográficos. Es de paga (y bastante cara).

QGIS (Quantum GIS): Es una herramienta SIG **de código abierto muy popular**. Ofrece muchas de las mismas funcionalidades que ArcGIS y es utilizada tanto por profesionales como por aficionados debido a su accesibilidad y flexibilidad (y que es gratis).



Google Earth Engine: Es una plataforma para el análisis de grandes conjuntos de datos geoespaciales. Es utilizada principalmente para el análisis y visualización de datos geográficos a escala global.

Herramientas de manejo de información geográfica.



AutoCAD Map 3D: Desarrollado por Autodesk, combina las capacidades de CAD con funcionalidades SIG, lo que permite a los usuarios trabajar con datos geoespaciales directamente en un entorno de diseño.

Otras más de nicho:

1. **GRASS GIS.** (Modelado)
2. **ERDAS IMAGINE** (Análisis de imágenes espaciales)
3. **POSTGIS** (Extensión de la base de datos PostgreSQL para almacenar y gestionar datos espaciales.)



¿Por qué usar R?



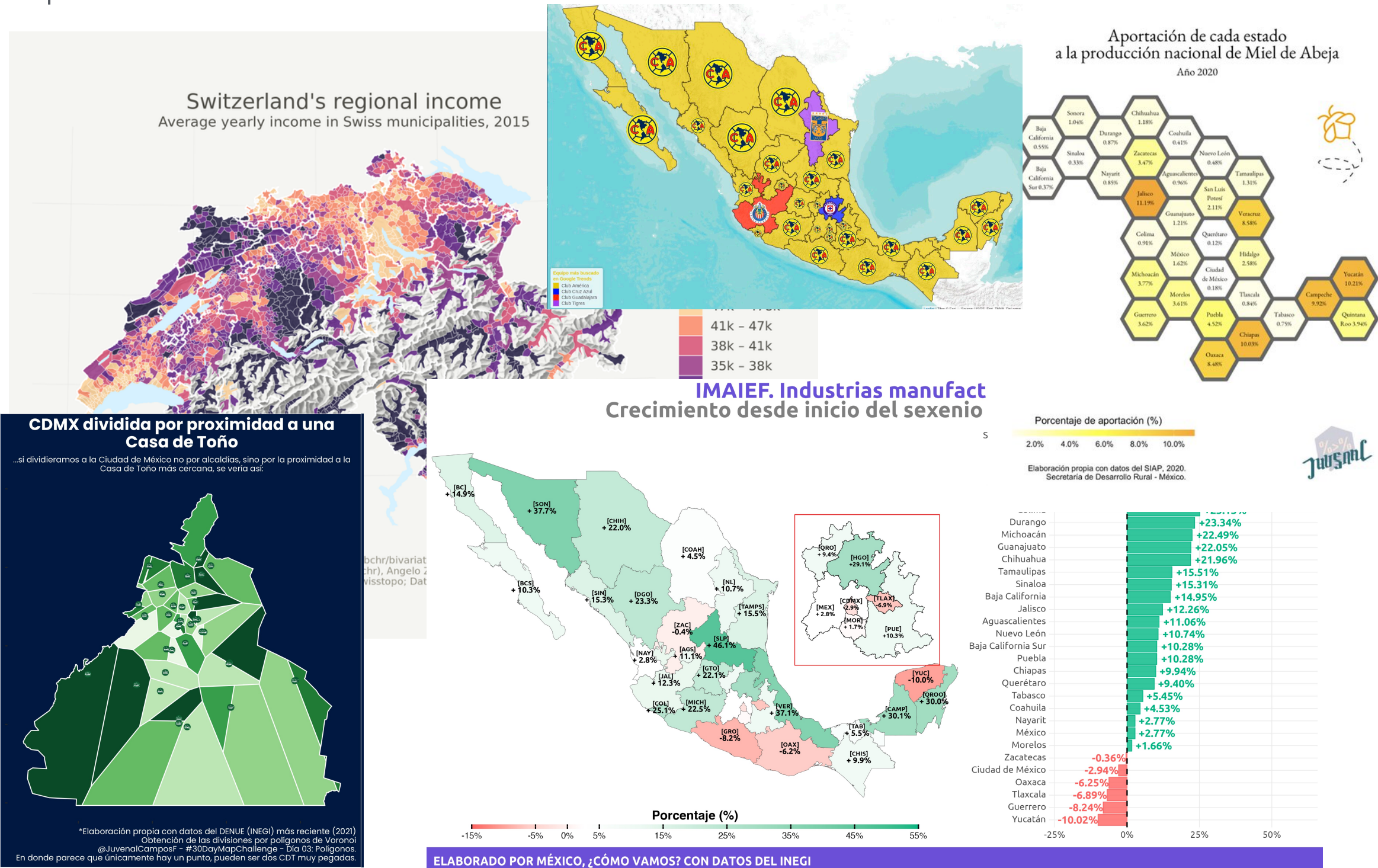
Algunas ventajas de usar R (o Python) para el análisis geoespacial:

- 1) **Integración.** Hacemos todo el análisis de un proyecto en la misma herramienta y el mismo espacio de trabajo.
- 2) **Automatización.** Es posible automatizar procesos complejos y repetitivos usando scripts.
- 3) **Comunidad y recursos.** Hay una gran cantidad de paquetes, bibliotecas y una comunidad activa que comparte recursos y ayuda.

¿Por qué usar R?



La **gran mayoría** de los mapas que se usan en economía se pueden hacer en R






Información geográfica


Información geográfica



La información geográfica es aquella información que tiene algún componente espacial, es decir, una **ubicación**, y además, una información **atributiva** que nos detalle más sobre ese elemento en cuestión.

Info geográfica: **componente espacial** + **atributo(s)**

- 
- Puntos
 - Líneas
 - Polígonos
 - Áreas
 - Matrices de píxeles

- 
- Nombres de calles
 - Estadísticas de un lugar
 - Precios
 - Población
 - Índices

Archivos para guardar información geográfica:



La información geográfica se almacena en archivos especiales. Los más comunes son los siguientes:



ESRI Shapefile.

Es el mas común, diseñado para funcionar con ArcGIS y QGis. Está conformado por 5 archivos diferentes que almacenan geometrías, atributos y metadatos.



{JSON, GeoJSON, TopoJSON}

Uno de los más populares, actualmente entre gente que usa R o Python. Es ligero, fácil de almacenar y de acceder desde la web.



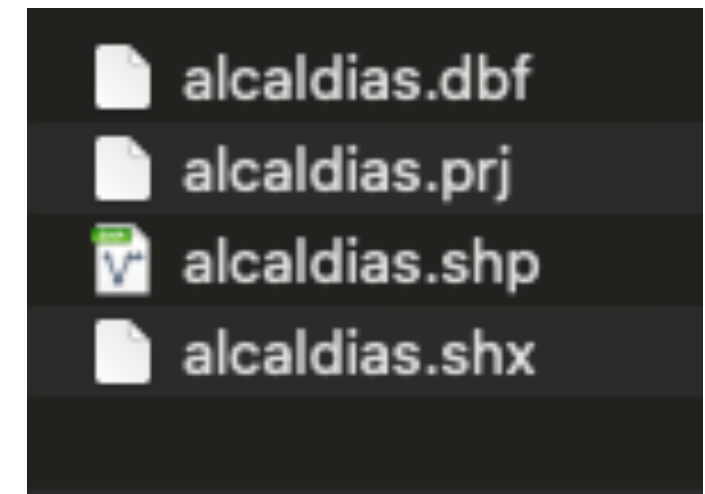
Google *.kml*.kmz

Es el formato diseñado por Google para trabajar utilizando las herramientas de Google Maps.

Archivos para guardar información geográfica:



Ojo 👁️: Un *shapefile* generalmente está **compuesto de múltiples archivos**, no solo del archivo *.shp. Si se borra uno de los archivos componentes es posible que el shapefile ya no cargue bien :(.

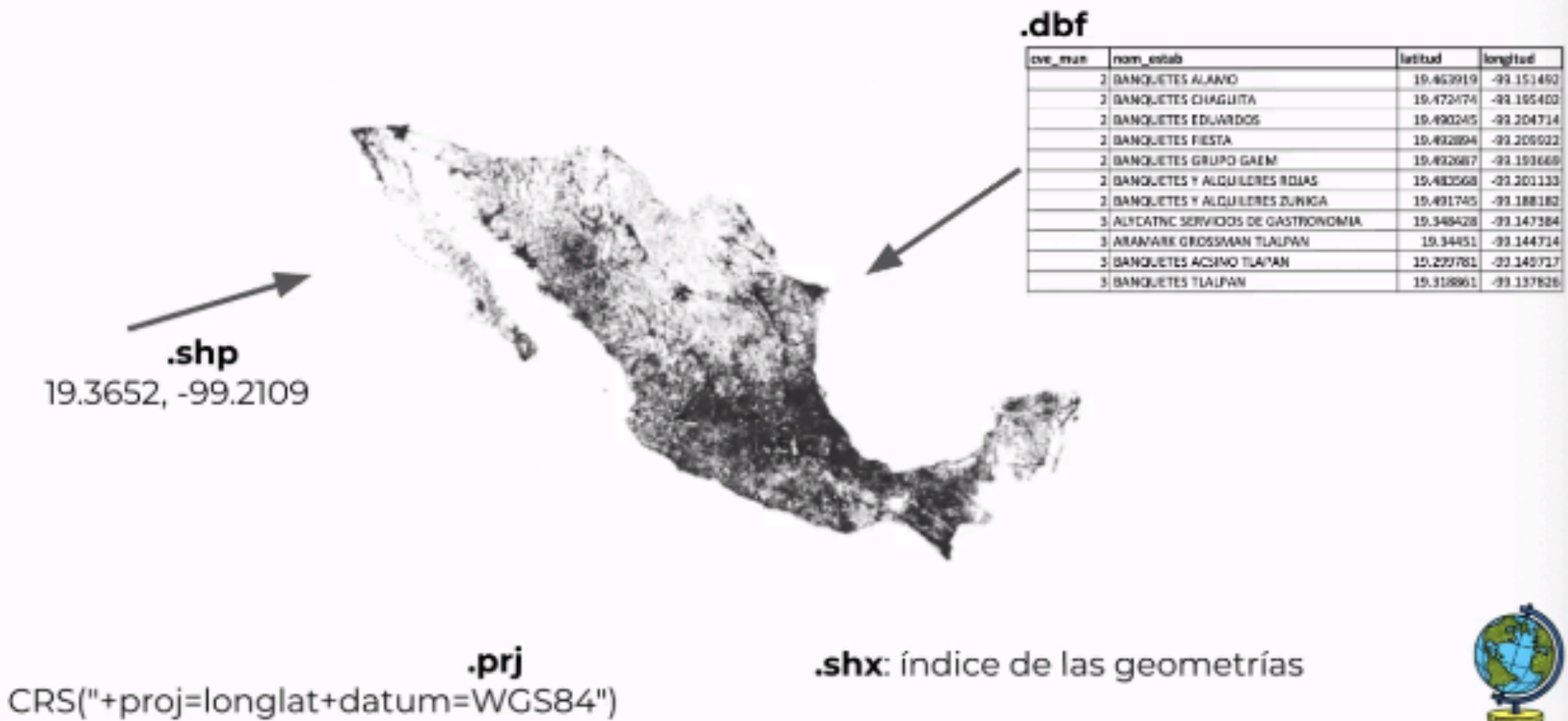


- Los archivos principales de un shapefile son:
 - **.shp**: almacena la geometría (forma) de las entidades.
 - **.dbf**: es la base de datos en formato dBase, almacena los atributos de las entidades.
 - **.shx**: funciona como índice de la geometría de las entidades.
 - **.prj** contiene información sobre el sistema de coordenadas o **proyección cartográfica**.
- Otros archivos que puedes encontrar:
.mxs, .cpj, .ixs, .fbn, .sbn, entre otros.



Archivos para guardar información geográfica:

COMPONENTES SHAPEFILE



Archivos para guardar información geográfica:



La información geográfica se almacena en archivos especiales. Los más comunes son los siguientes:



Una hoja de excel con coordenadas (.csv, .xlsx)

Una hoja de excel también puede almacenar información geográfica, especialmente si almacena coordenadas de puntos o ubicaciones.



Imagen .TIFF, .GEOTIFF
Formato más utilizado para almacenar bases de datos Raster e imágenes satelitales. En términos prácticos, es una imagen grande y georreferenciada.



Conceptos importantes
(antes de hacer mapas)



Mapa



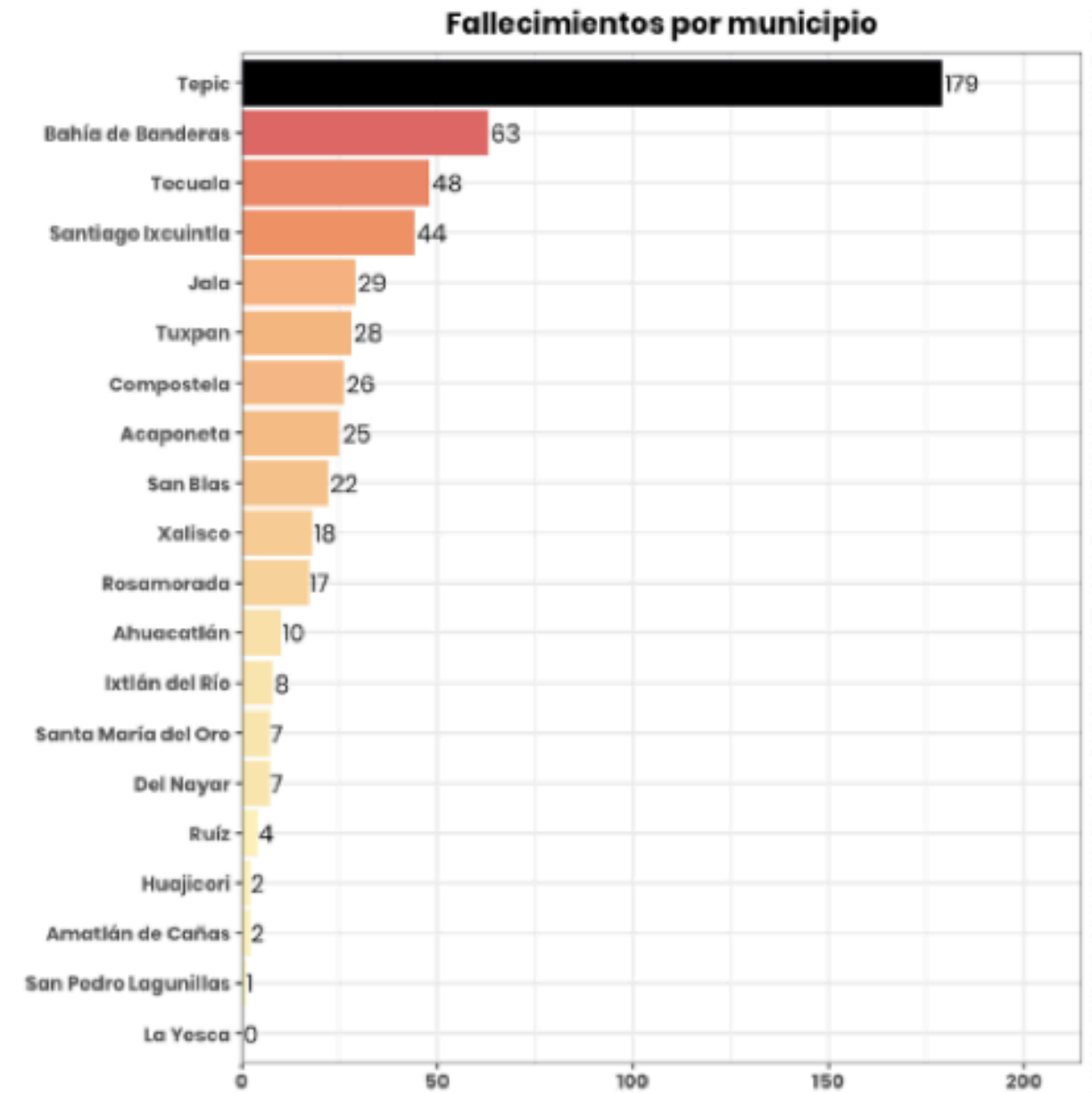
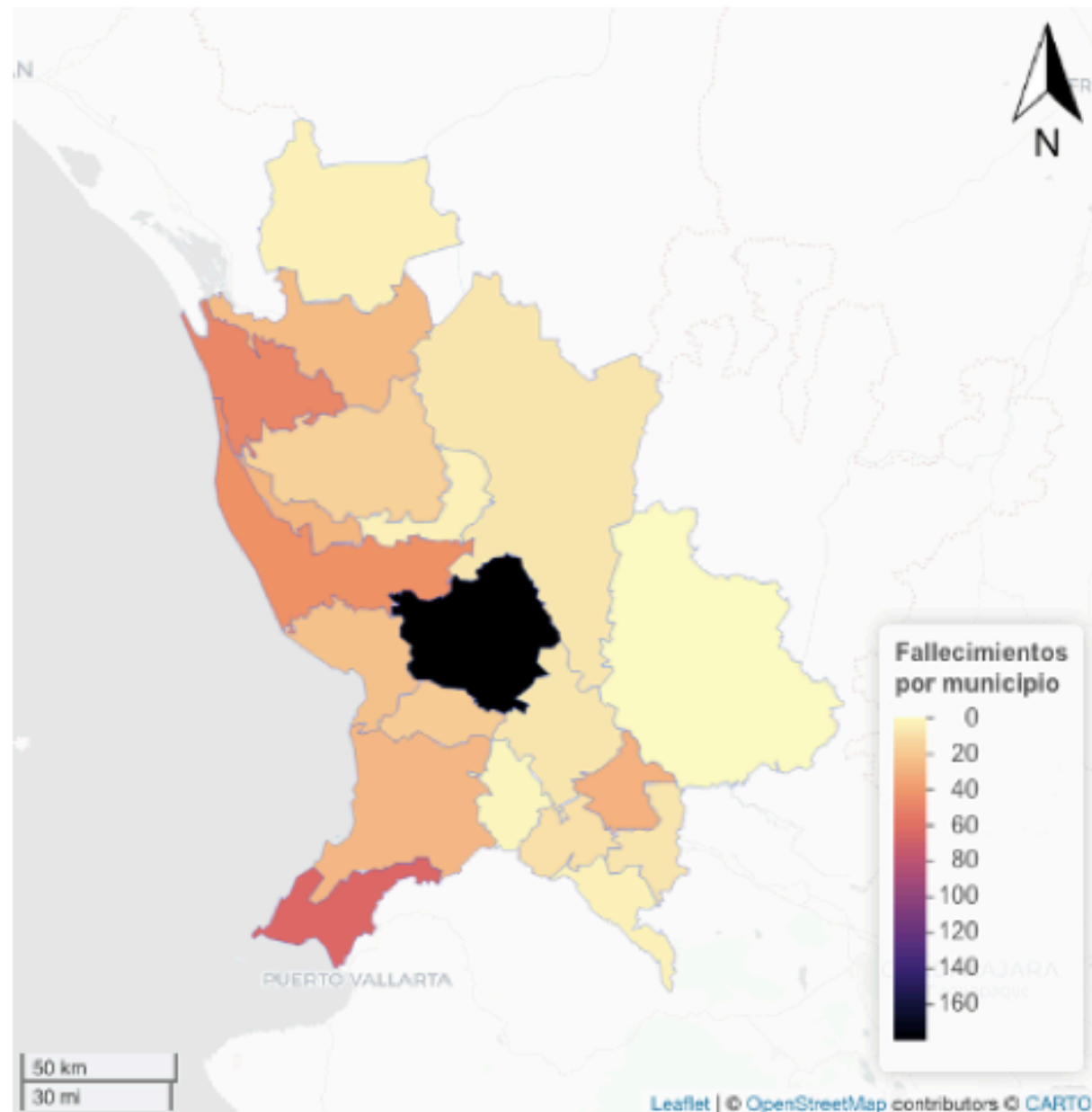
- Un mapa es una **representación gráfica simplificada de un territorio** con propiedades métricas sobre una superficie bidimensional (Wikipedia, 2020).
- Un mapa sirve para poder visualizar información que varía o se distribuye a lo largo del espacio.

Si bien un mapa es una muy buena alternativa para visualizar algún fenómeno o variable, siempre hay que tener en cuenta el propósito al elaborar uno de estos.

Uso de mapas



¿Cuál es **mejor**? ¿Qué **variable** y **mensaje** da cada una?

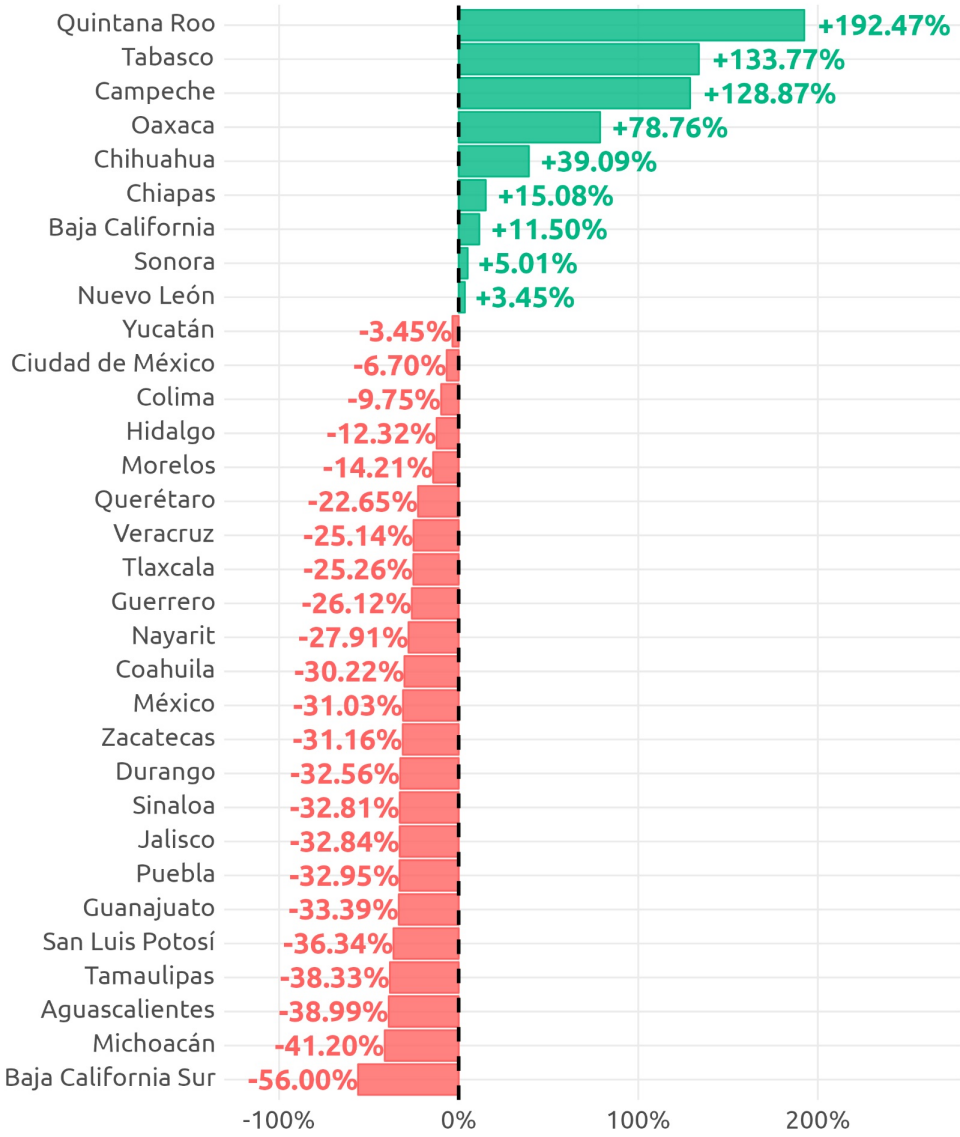
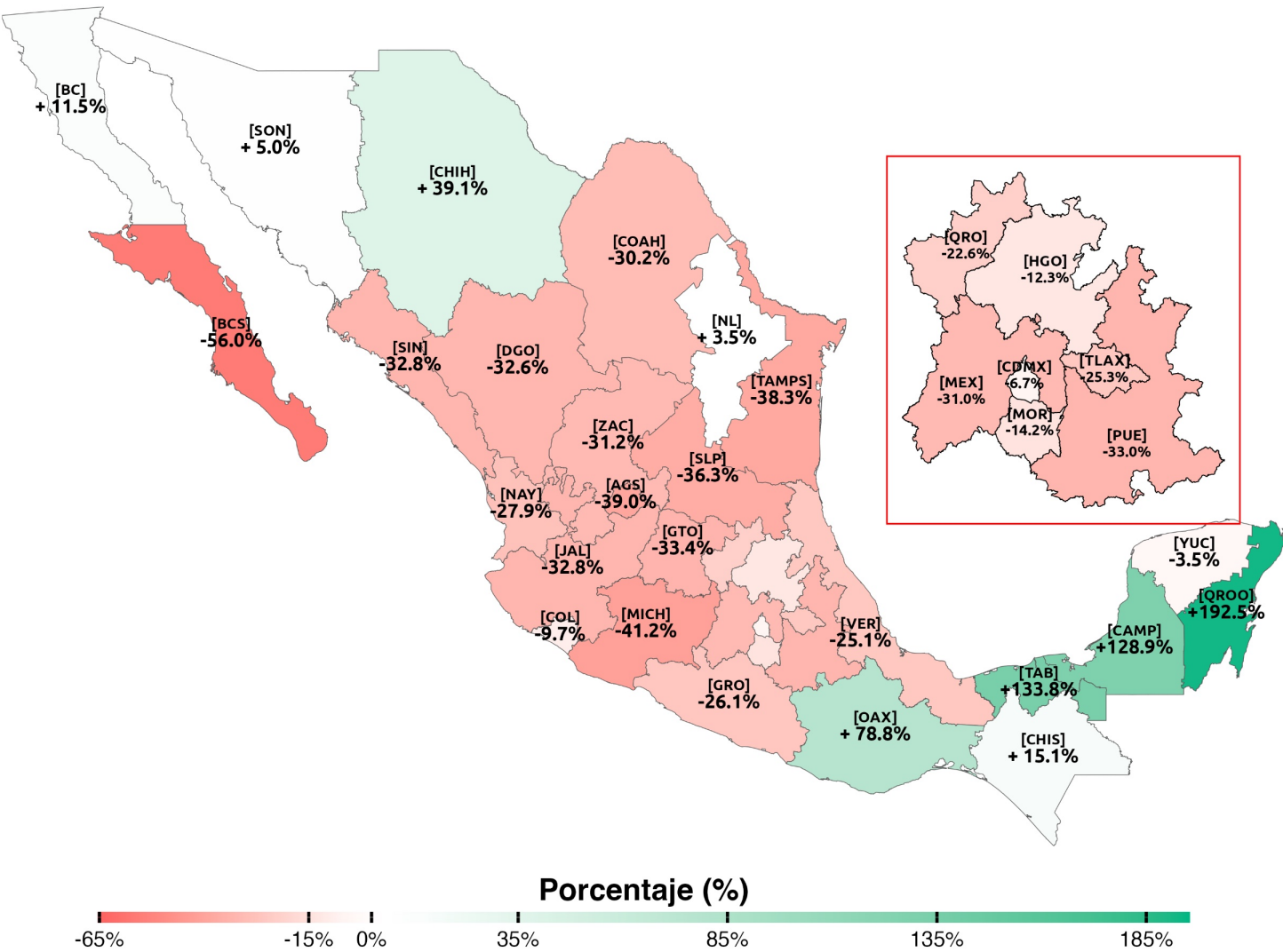


Fuente: Elaboración propia con datos de la Secretaría de Salud, 24 de Agosto del 2020.

Uso de mapas



IMAIEF. Construcción Crecimiento desde inicio del sexenio a Febrero 2024



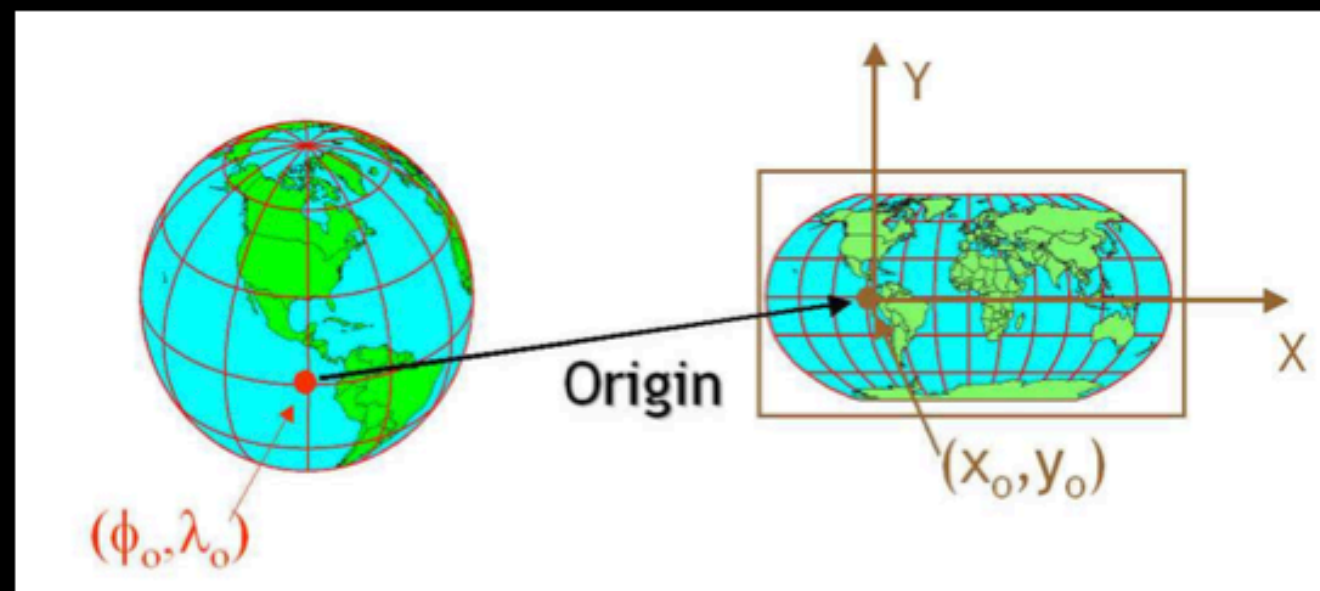
Sistema de coordenadas de referencia



Un CRS es el **sistema de coordenadas** que se utiliza para **localizar** las entidades geográficas.

A grandes rasgos, un crs debe incluir: **a) una proyección geográfica, b) un punto de referencia c) un sistema de traducción a otro crs** y d) un datum o elipsoide de referencia.

En esta clase utilizaremos el **crs 4326** (WGS84 proyección lat/long) por ser el **más común** y el que utilizan por default librerías como leaflet y google maps.

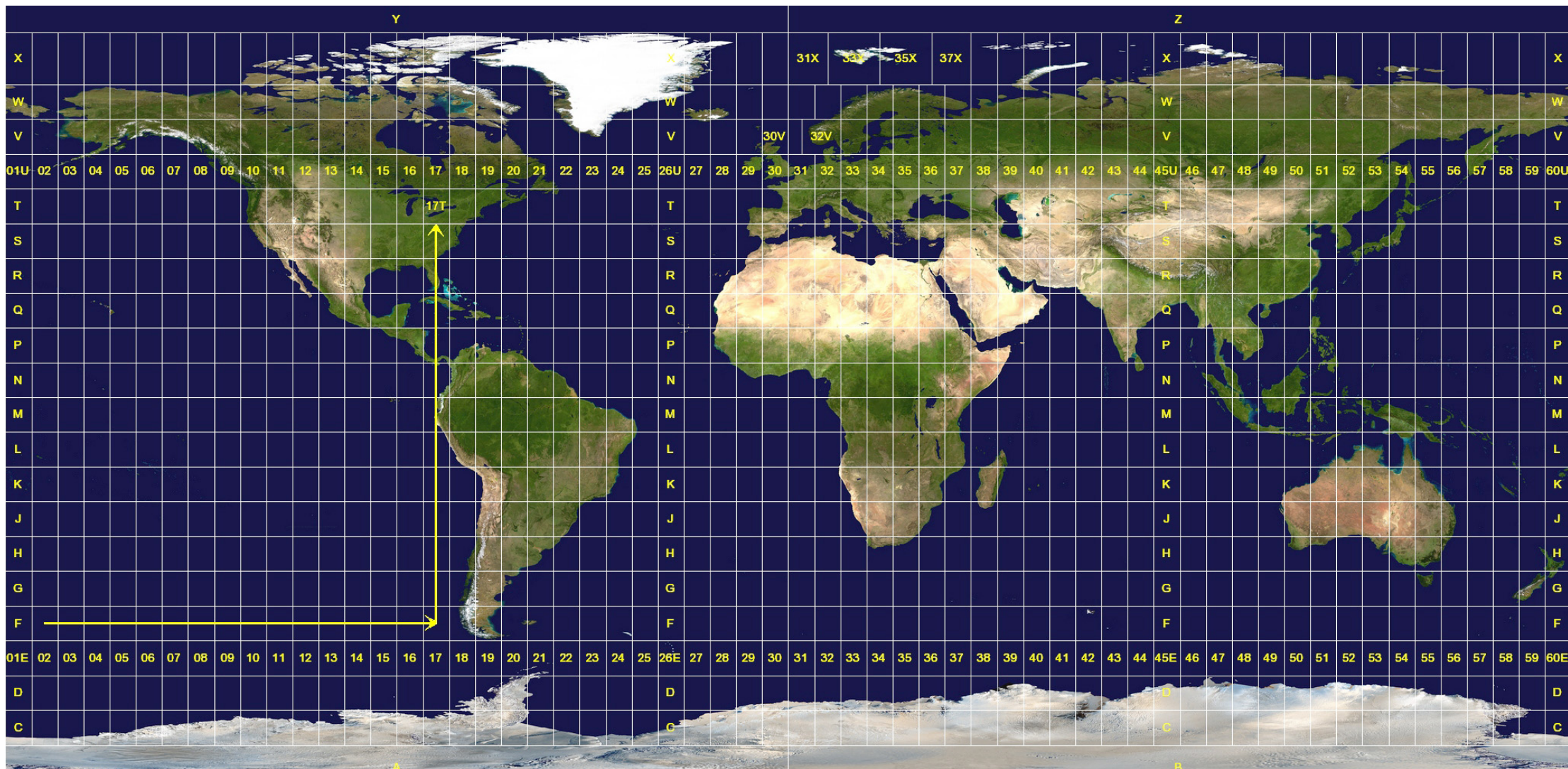


Sistema de coordenadas de referencia



Otros sistemas de coordenadas de referencia útiles:

- **EPSG:6372**, proyección default de INEGI. Tomar en cuenta que los datos del INEGI pueden no siempre venir en lat/lon.
- **EPSG:2163**: para mapas de EEUU (US National Atlas Equal Area)
- **EPSG:6362** (UTM zone 14N): Para pasar de lat/lon a metros dentro de la zona 14 de Mercator (CDMX).



Claves INEGI



Las claves Geoestadísticas de INEGI son números identifican a un estado, municipio, AGEB, comunidad, colonia o manzana en particular.

Estas claves funcionan como **el Id de un polígono**, y son **muy importantes para utilizarlas como variables "Llave"** para pegar bases de datos de atributos con bases de datos geográficas.

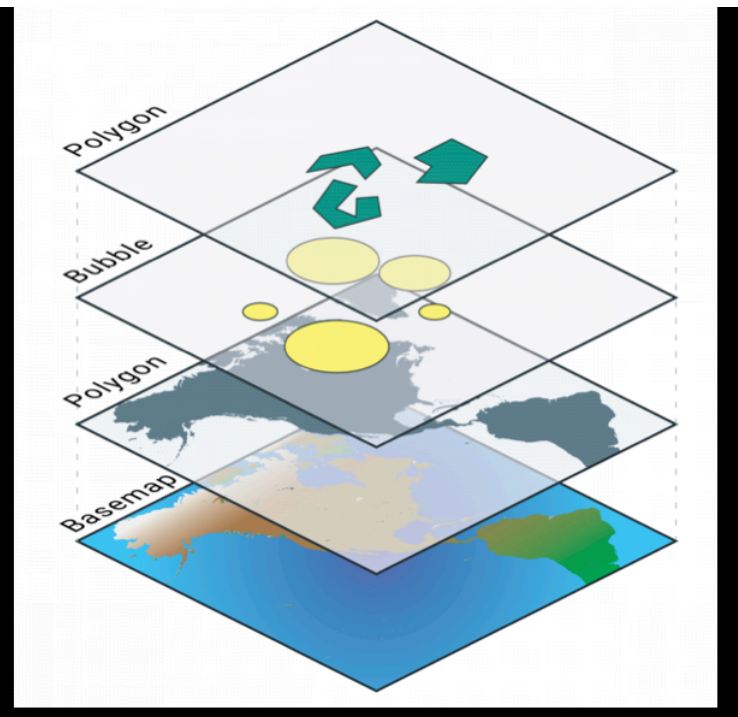
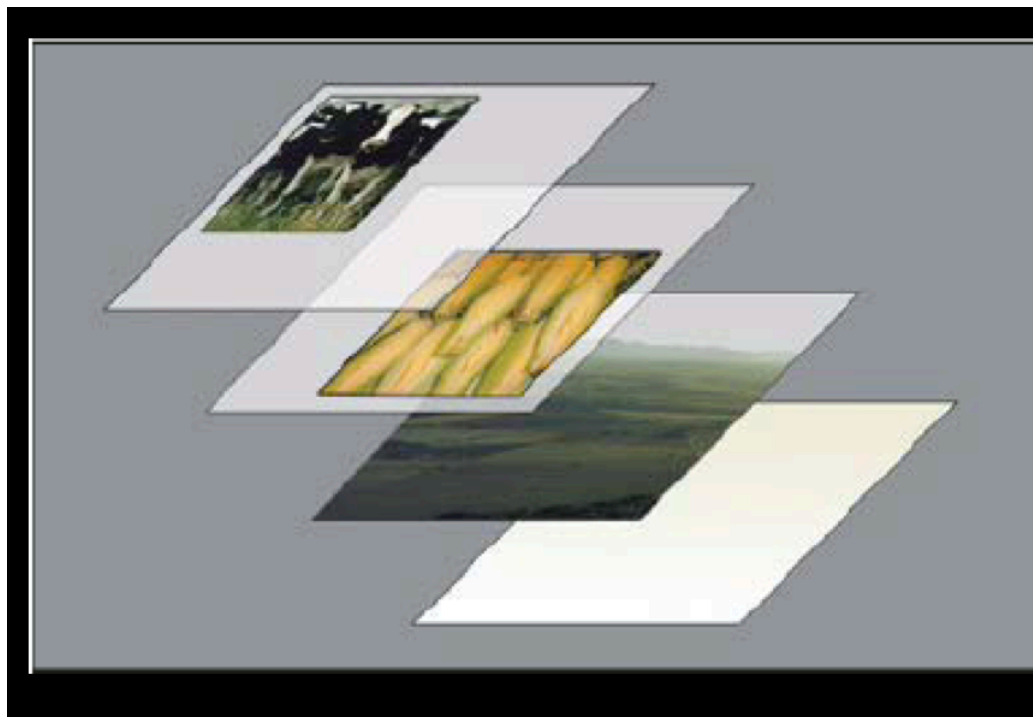
	CVEGEO	CVE_ENT	CVE_MUN	NOM_MUN
1	01001	01	001	Aguascalientes
2	01002	01	002	Asientos
3	01003	01	003	Calvillo
4	01004	01	004	Cosío

Capas



Las capas son una herramienta de trabajo que se asemeja a las hojas apiladas de acetato.

Las capas se modifican de manera individual, y al final se juntan (como un acetato) de tal forma que generen una visualización final.





Fuentes de información geográfica



Tiene información oficial de los límites estatales y municipales e información de referencia para preparar censos y otros productos del INEGI.

Marco Geoestadístico

El Marco Geoestadístico es un sistema único y de carácter nacional diseñado por el INEGI, el cual presenta la división del territorio nacional en diferentes niveles de desagregación para referir geográficamente la información estadística de los censos y encuestas institucionales y de las Unidades del Estado, que se integra al Sistema Nacional de Información

 <https://www.inegi.org.mx/temas/mg/>

Topografía

Información que tiene como objetivo representar la infraestructura, orografía, hidrografía y las poblaciones del país. En ella se registran fielmente todos estos elementos y las relaciones que guardan entre sí. Contiene información sobre el relieve, rasgos hidrográficos, vegetación densa, áreas con actividades agrícolas, localidades y vías de comunicación, además de

 <https://www.inegi.org.mx/temas/topografia/>

Descarga masiva

En esta sección podrás descargar los archivos del DENU, el Banco de Indicadores, el Inventario Nacional de Viviendas, Microdatos y la Sala de prensa por área geográfica, proyectos, años de la información, temas y

 <https://www.inegi.org.mx/app/descarga/>



Otras fuentes



2.2 Datos de la CONABIO.

La CONABIO Posee un gran inventario de información geográfica que permite acceder a múltiples capas de información tanto de interés biológico y físico como de interés social.

Para acceder a estas capas podemos ingresar al siguiente enlace:

Geoportal del Sistema Nacional de Información sobre Biodiversidad [6,660 mapas] - CONABIO

El Geoportal de la CONABIO tiene como objetivo facilitar la localización, consulta y obtención de la cartografía temática generada y recopilada por la Comisión. La planeación, el desarrollo y la implementación de nuestro portal se realizó con software libre y de código abierto.


 <http://www.conabio.gob.mx/informacion/gis/>

2.3 Datos de los Atlas de Riesgo y de gobiernos locales.

El atlas de riesgo Nacional es una buena referencia para descargar igualmente capas de información geográfica, ya que tiene un montón.

Atlas Nacional de Riesgos

Sistema de Información sobre Riesgos El sistema integra todos los mapas del Atlas Nacional de Riesgos, de peligro, exposición, vulnerabilidad y riesgo, clasificados según el tema para su visualización y análisis. continuar leyendo

 <http://www.atlasnacionalderiesgos.gob.mx>



Otras fuentes




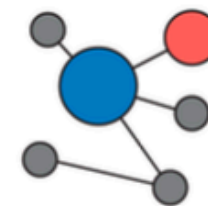
UN - The humanitarian data exchange.

Plataforma abierta para compartir datos de potencial uso humanitario entre organizaciones, facilitando su uso y localización para realizar multitud de análisis.

Search for a Dataset - Humanitarian Data Exchange

Contains data from the World Bank's data portal. There is also a consolidated country dataset on HDX. The world economy needs ever-increasing amounts of energy to sustain economic growth, raise living standards, and reduce

 https://data.humdata.org/search?q=mexico&ext_search_source=main-nav



HDX
Humanitarian
Data Exchange

GeoComunes

Colectivo que trabaja acompañando a los pueblos, comunidades, barrios, colonias u organizaciones de base que en la lucha por la defensa de los bienes comunes requieran de la producción de mapas para su análisis y difusión con la

<http://geocomunes.org/>

A stylized, hand-drawn illustration of an open map. The map is divided into several panels by black outlines. The colors used are light blue for water and green for land. A red location pin with a yellow center is placed on the right side of the map. A dark red banner with white text is overlaid on the map.

Trabajando información geográfica en R

Haciendo mapas en R

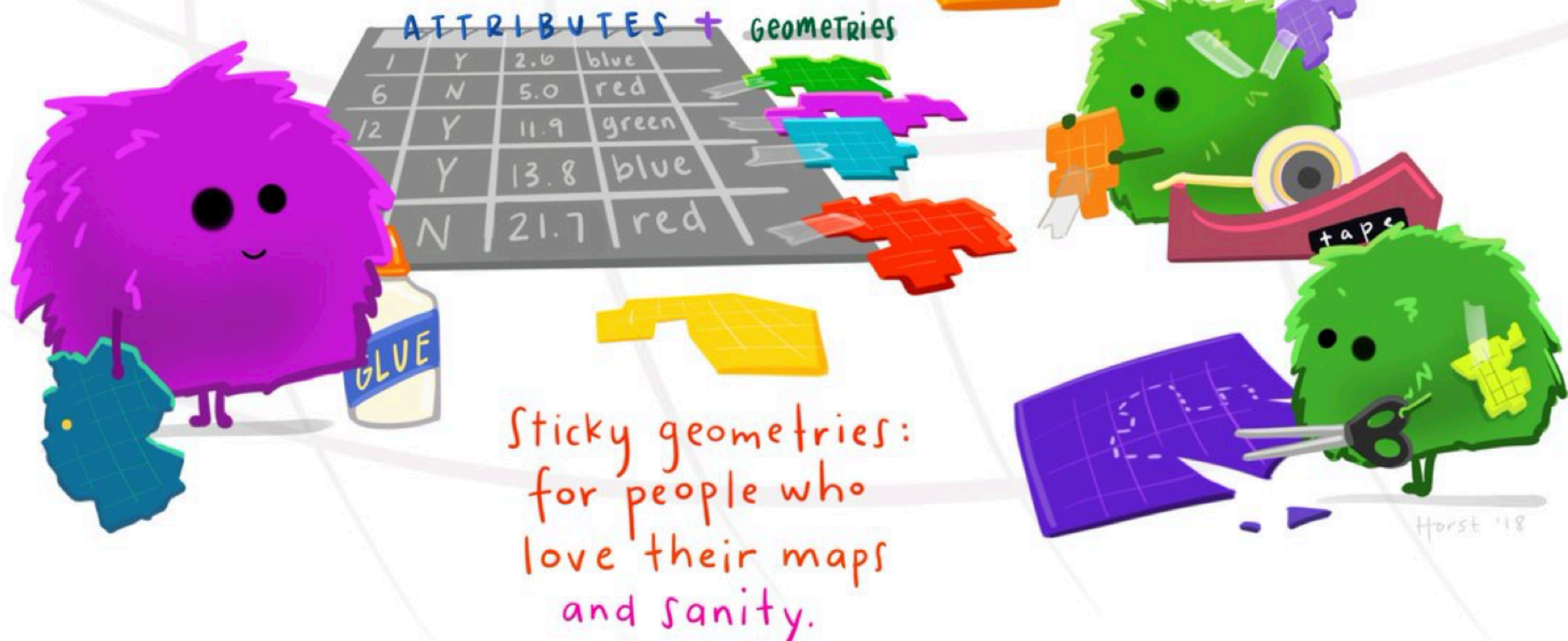


- Para hacer mapas en R, necesitamos lo siguiente:
1. Un **archivo que almacene la información geográfica** (lo que vimos en la anterior actividad)
 2. Una **librería para abrir ese archivo** {sf}
 3. Una **base de atributos** (una tabla de datos)
 4. Una **librería para manejar esos atributos** {dplyr,tidyverse}
 5. Una **librería para visualizar esa información** {ggplot, leaflet}



=



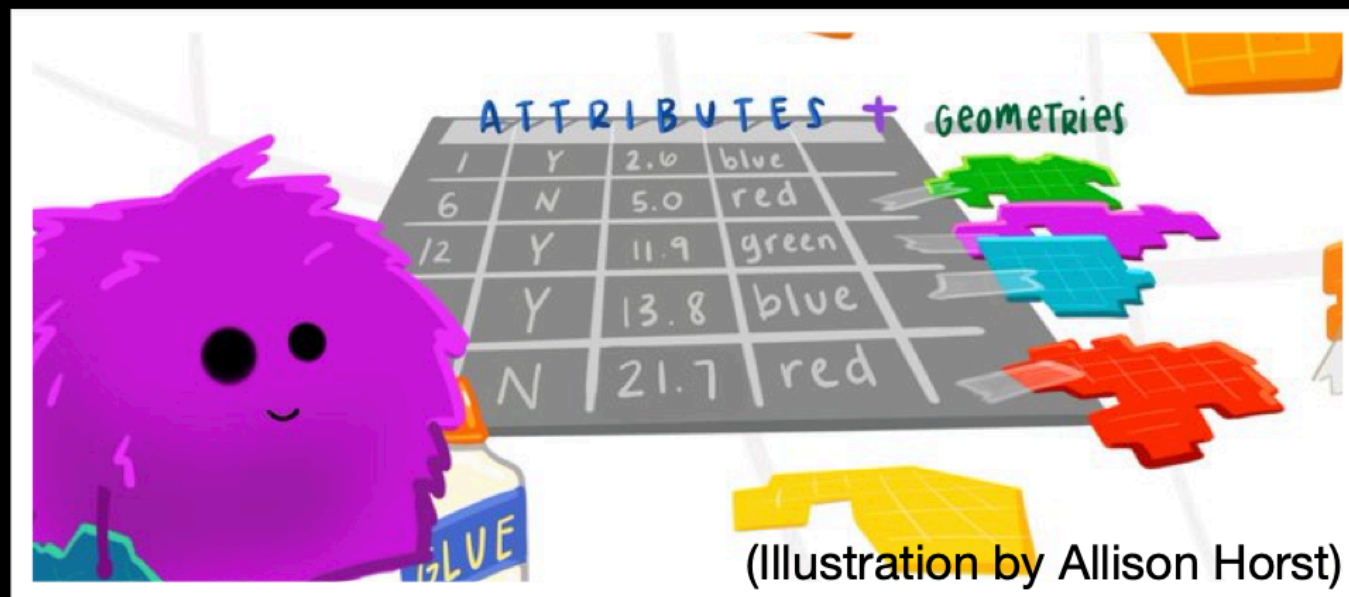


Librería {sf}



Con la librería sf los objetos espaciales son **almacenados como un data frame simple** con **una columna que contiene la información de las coordenadas de las geometrías.** (Objeto clase sf)

Esa columna especial es una lista de la misma longitud que el número de filas en el data frame.



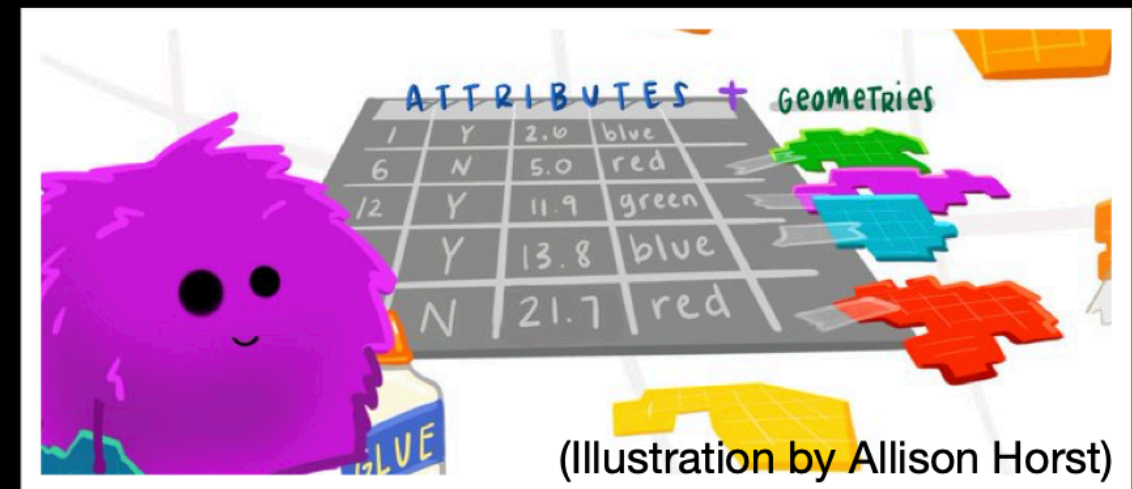
Función `sf::st_read()` o `sf::read_sf()`



La función `st_read()` lee un archivo o base de datos vectorial como un objeto `sf`.



1. `sf::st_read()`



Objeto clase `sf - data.frame`, que se puede manipular con los verbos `dplyr` que vimos con Segasi.

Argumentos: `dsn`: ubicación del archivo como cadena de texto. Acepta direcciones url, por lo que podemos leer archivos de internet.

Función `sf::st_read()` o `sf::read_sf()`



Ejemplo de uso ya en R. (Previamente hay que descargar la librería).

```
library(sf)
# 1. Podemos leer archivos de internet
mpios <- st_read("https://raw.githubusercontent.com/JuveCampos/mpios.geojson")
# 2. Podemos leer archivos localmente
mpios <- st_read("01_Datos/mpios.geojson")
# 3. Podemos leer shps o kml
alcaldias <- st_read("01_Datos/alcaldias.kml")
# Hay que tener todos los archivos para que funcione
frontera <- st_read("01_Datos/Shape Frontera/Mexico_and_US_Border.shp")
```

`plot(mpios, max.plot = 1)`



`plot(alcaldias, max.plot = 1)`



`plot(fronteras, max.plot = 1)`



Objeto *sf*



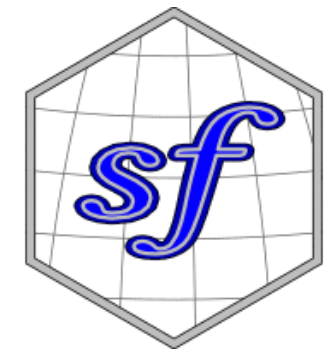
La clase *sf* es la clase de un **objeto** que almacena las coordenadas de los polígonos, líneas o puntos necesarios para crear un mapa.

Generalmente, viene acompañado de la clase *data.frame*, lo cual **permite que podamos utilizar en estos objetos los verbos *dplyr* para trabajar con la información** de las variables que acompañan a dicha geometría (con las funciones *filter*, *summarise*, *select*, *arrange*, *mutate*, etc.).

```
> class(mpios)
[1] "sf"          "data.frame"
```

Para verificar que el objeto sea clase *sf*, utilizamos la función *class()*.

Objeto *sf*



```
> mpios
```

```
Simple feature collection with 2465 features and 3 fields
```

```
geometry type:  MULTIPOLYGON
```

```
dimension:      XY
```

```
bbox:           xmin: -118.3651 ymin: 14.5321 xmax: -86.71041 ymax: 32.71865
```

```
geographic CRS: WGS 84
```

```
First 10 features:
```

	CVEGEO	NOM_ENT	NOM_MUN	geometry
1	01001	Aguascalientes	Aguascalientes	MULTIPOLYGON (((-102.1064 2...
2	01002	Aguascalientes	Asientos	MULTIPOLYGON (((-102.0519 2...
3	01003	Aguascalientes	Calvillo	MULTIPOLYGON (((-102.6857 2...
4	01004	Aguascalientes	Cosío	MULTIPOLYGON (((-102.2879 2...
5	01005	Aguascalientes	Jesús María	MULTIPOLYGON (((-102.3357 2...
6	01006	Aguascalientes	Pabellón de Arteaga	MULTIPOLYGON (((-102.2535 2...
7	01007	Aguascalientes	Rincón de Romos	MULTIPOLYGON (((-102.2268 2...
8	01008	Aguascalientes	San José de Gracia	MULTIPOLYGON (((-102.4561 2...
9	01009	Aguascalientes	Tepezalá	MULTIPOLYGON (((-102.1799 2...
10	01010	Aguascalientes	El Llano	MULTIPOLYGON (((-102.0295 2...

Información de la geometría

- Incluye el número de filas (features) y columnas (fields)
- El tipo de la geometría,
- las dimensiones (XY y Z)
- Las coordenadas extremas (bbox)
- El Sistema de coordenadas de referencia

Atributos:

- Incluye las variables asociadas a cada una de las geometrías.

Geometrías:

- Incluye el tipo de la geometría y, en este caso, los vértices de los polígonos de los municipios.

Funciones de sf:



Función `sf::st_crs()`

La función `sf::st_crs()` nos indica en qué sistema de coordenadas de referencia está nuestro objeto `sf`.

```
st_crs(mpios)
# > Coordinate Reference System:
#   User input: WGS 84
```

```
st_crs(alcaldias)
# > Coordinate Reference System:
#   User input: WGS 84
```

```
st_crs(frontera)
# > Coordinate Reference System:
#   User input: WGS 84
```


Funciones de sf:



Función `sf::st_transform()`

La función `st_transform()` nos permite transformar el `crs` a otro `crs` que nos convenga más (*generalmente, el `crs` = 4326*).

Ejemplo de uso:

```
encharcamientos_2018 <- st_read("http://www.atlas.cdmx.gob.mx/datosAbiertos/SPC_Encharcamientos_2018.geojson")
st_crs(encharcamientos_2018)
# Coordinate Reference System:
#   User input: Mexico ITRF2008
#   ID["EPSG",6365]]

# Si lo quiero transformar a 4326:
encharcamientos_2018 <- encharcamientos_2018 %>%
  st_transform(crs = 4326)
```

```
> encharcamientos_2018
Simple feature collection with 121 features and 20 fields
geometry type:  POINT
dimension:      XY
bbox:           xmin: -99.24384 ymin: 19.19662 xmax: -98.95801 ymax: 19.53632
geographic CRS: WGS 84
```

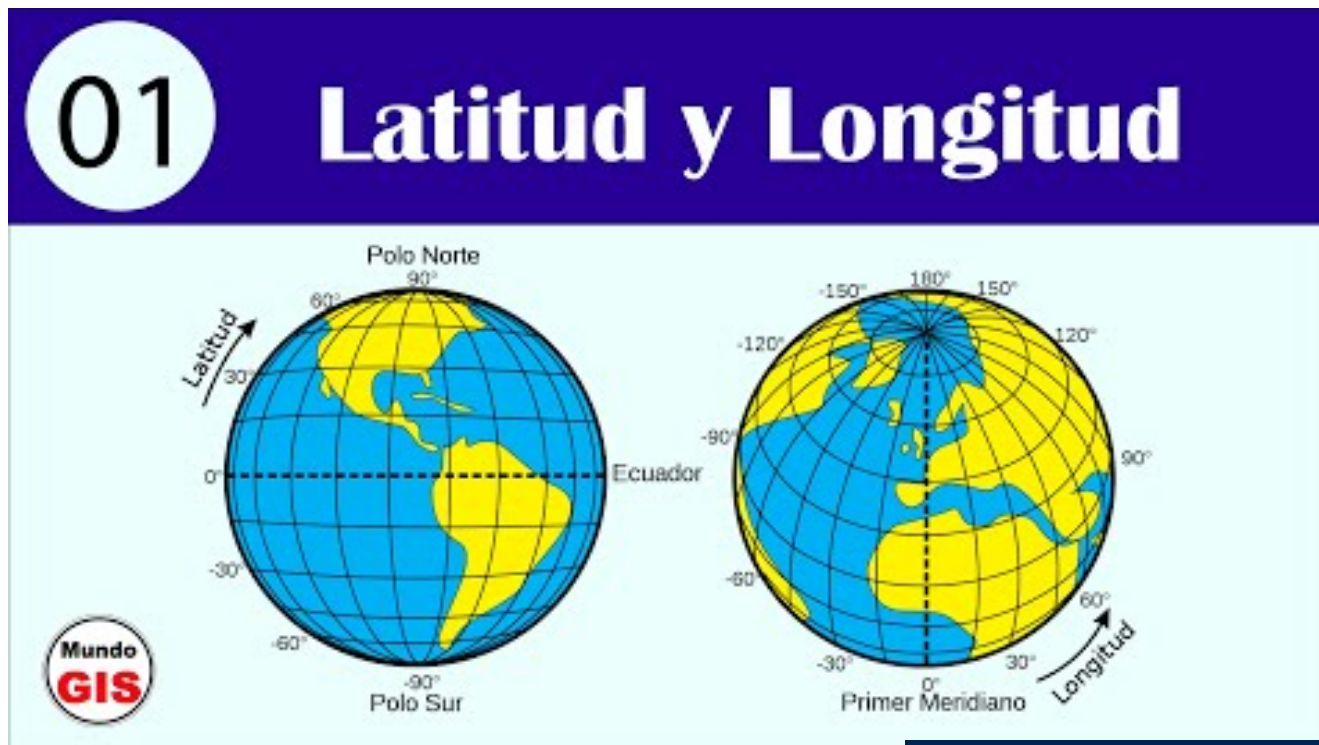


Funciones de sf:



Función `sf::st_as_sf()`

Permite transformar un objeto del tipo *data.frame* en un objeto *sf*, si cuenta con datos de coordenadas. Esta función se usa comúnmente para tratar con datos de **puntos** que vienen desde un excel.



```
delitos <- readxl::read_xlsx("01_datos/delitos_2019.xlsx")
delitos_shp <- st_as_sf(delitos,
                        coords = c("longitud",
                                   "latitud"),
                        crs = 4326)
```



Visualización

Función *plot()*



La función **plot()** nos permite hacer un mapa exploratorio de las geometrías que contiene el objeto clase sf,

```
plot(mpios, max.plot = 1)
```



```
plot(alcaldias, max.plot = 1)
```



```
plot(fronteras, max.plot = 1)
```



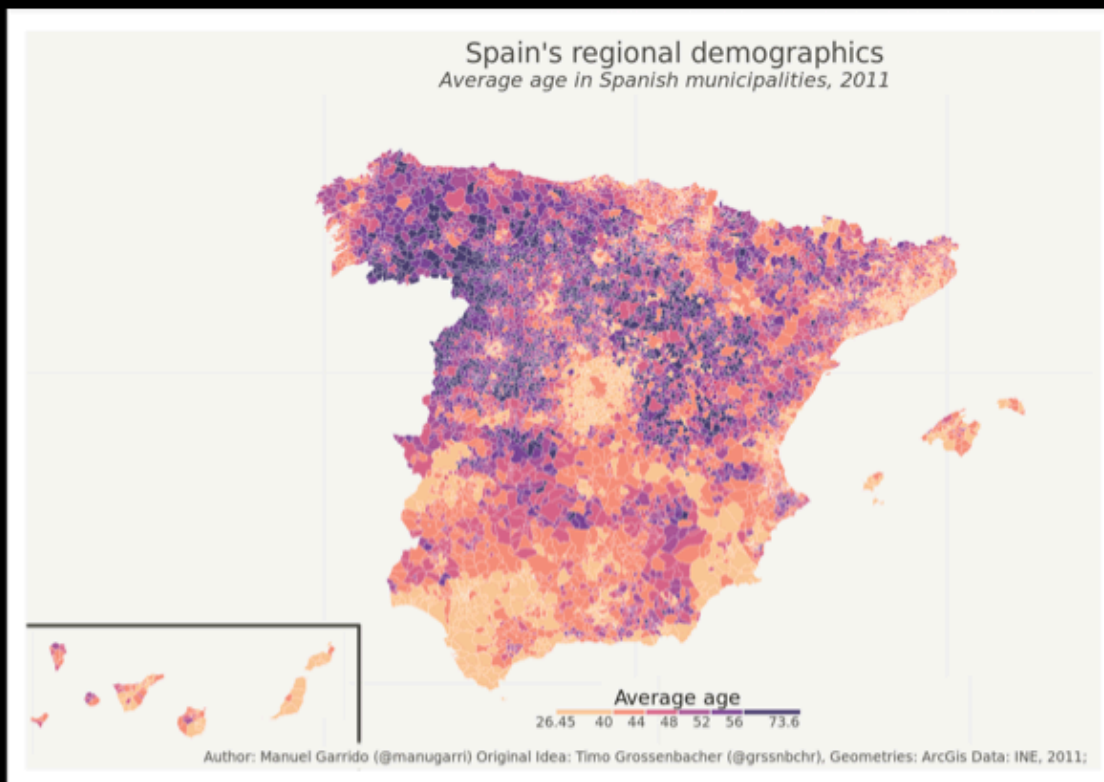
El argumento `max.plot = 1` imprime solo un mapa de la primera variable. Si no se incluye este argumento, imprimirá 8/9 mapas de las primeras variables (atributos) del objeto sf, **y esto puede congelar la computadora**, si el objeto es muy pesado.

Visualización



La visualización de mapas se puede hacer de dos maneras: de manera estática `{ggplot2}` e interactiva `{leaflet}`.

`{ggplot2}`



Autor: Manuel Garrido @manugarri

`{leaflet}`



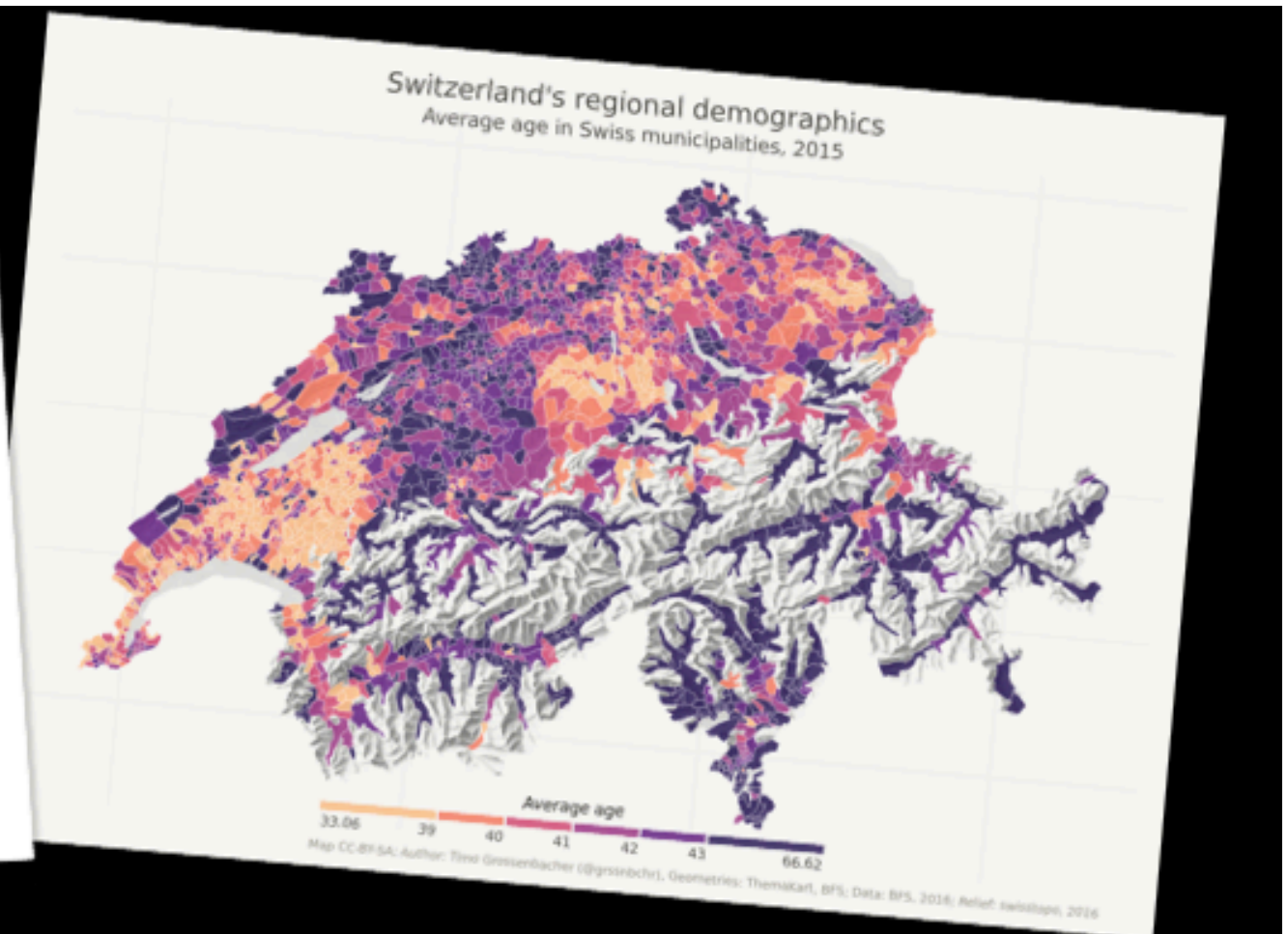
Autor: Charles Murray. Ver enlace para más referencias.

Librería `{ggplot2}`



La función **`geom_sf()`** es la función que nos va a permitir añadir la capa del elemento geométrico en una gráfica ggplot.

Es similar a `geom_histogram()`, `geom_point()` o `geom_line()`, solo que esta forzosamente requiere un objeto clase `sf` para funcionar.



Librería *{ggplot2}*



Ejemplo de uso:

```
# Hacemos el mapa
mapx <- map_estado %>%
  # Declaramos los elementos estéticos
  ggplot(aes(fill = Valor)) +
  # Generamos el mapa (y ponemos los perimetros en blanco)
  geom_sf(color = "gray90") +
  # Cambiamos los colores del relleno
  scale_fill_gradientn(colors = viridis(begin = 0,
                                         end = 1,
                                         n = 10))|
```



Sección práctica



=



Sección práctica



Ejercicios:

0. Descargue las librerías *sf*, *raster*, *cartogram* y *tidyverse*, en caso de que aún no las tenga instaladas.
1. Cargue el excel de Delitos 2019 y, a partir de las coordenadas de latitud y longitud, conviertan en un objeto *sf*. Grafique.
2. Cargue los polígonos de las alcaldías (*alcaldias.kml*). Explore los datos, identifique geometrías y atributos. Grafique
3. Cargue los datos de “municipios_2022_simplificado.rds”. ¿Qué información encuentra?
4. Cargue los datos del archivo “prec_9_extrema.tif” sobre precipitación extrema en territorio mexicano. ¿Qué tipo de información es? Grafique.

Sección práctica



Ejercicios:

5. Exporte solo los delitos pertenecientes a la alcaldía Cuauhtémoc a un archivo llamado "delitos_cuauhtemoc.geojson".
6. Cargue el archivo de "Índice de Desarrollo Humano.csv" y replique en ggplot el siguiente mapa (sin el post-it).

