

Visualización y geolocalización de datos

Diego Hernangómez Gema Fernández-Avilés

2022-02-01

Índice general

Introducción	2
1. La revolución de los geodatos	3
2. Datos geográficos	5
2.1. Contexto general	5
2.2. Conceptos clave	8
3. Formatos de datos espaciales	10
3.1. Ficheros de datos espaciales	10
3.2. Modelo de datos	12
3.3. Sistema de Referencia de Coordenadas (CRS)	19
4. Estadística espacial	34
4.1. Antes de continuar... dependencia espacial.	35
4.2. Datos espaciales	37
4.3. Clasificación de datos espaciales	39
A. Tipos de CRS proyectados	42
A.1. Por tipo de superficie de proyección	42
A.2. Por métrica a preservar	44

Introducción



Objetivos de aprendizaje

¿Por dónde empezamos? Recursos interesantes

Libros de referencia:

- [Spatial Data Science with applications in R](#)
- [Geocomputation with R](#)
- [Displaying time series, spatial and space-time data with R](#)

Recursos de estadística espacial en R:

- [rspatial](#)
- [R-spatial](#)

Capítulo 1

La revolución de los geodatos

Que estamos en la la *era del dato*, que los *datos son el petróleo del siglo XXI* y que estamos rodeados de datos es una cuestión que ya hemos hecho inherente a nosotros. Vivimos en el momento del dato, donde la profesión de *Data Scientist* se ha convertido en la **profesión más sexy del siglo XXI** según vaticinó en 2012 [Harvard Business Review](#). Cada segundo se producen 1,7 MB de datos/persona y cada año esta cifra se duplica se duplica.

Este incremento exponencial de los datos ha sido posible, sin duda, gracias al desarrollo de la tecnología, la informática, los ordenadores, los teléfonos móviles, los satélites, internet, etc... y asociado a estas nuevas herramientas, se ha producido una lluvia sin precedentes hasta el momento de **datos espaciales o datos georreferenciados**. Cada teléfono inteligente tiene un **Receptor de posicionamiento Global** (en inglés, *Global Positioning System, GPS*) y, además, convivimos con una multitud de sensores en dispositivos que van desde satélites y vehículos semi-autónomos hasta científicos ciudadanos que miden incesantemente cada parte del mundo. La tasa de datos producidos es abrumadora. Un vehículo autónomo, por ejemplo, puede generar 100 GB de datos por día (The Economist, 2016).

Esta **revolución de los geodatos** y el **análisis de los datos espaciales** junto con los **Sistemas de Información Geográficos** (habitualmente expresados como **GIS** por las siglas de su nombre en inglés *Geographical Information System*) no sólo han impulsado la demanda de hardware informático de alto rendimiento y software escalable y eficiente para manejar y extraer la información, lo que se conoce como **Geocomputación**, sino que han dado lugar una nueva rama de conocimiento, **Ciencia de Datos Espaciales** comúnmente conocida como *Spatial Data Science* (SDS).

Como ejemplo de este abrumador desarrollo de datos georreferenciados, el Ministerio de Transportes, Movilidad y Agenda Urbana llevó a cabo durante los años 2020 y 2021 el denominado [Estudio de movilidad con Big Data](#), cuya fuente principal de datos fue el posicionamiento de los teléfonos móviles anonimizado. Estos datos permiten, por ejemplo, analizar la movilidad entre diversas zonas del territorio español de manera diaria (véase la Fig. 1.1). Este tipo de análisis era impensable hace tan solo unos años.

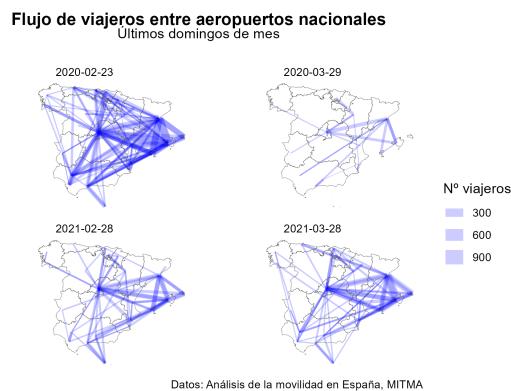


Figura 1.1: Análisis de movilidad COVID

Capítulo 2

Datos geográficos

2.1. Contexto general

La palabra geográfico puede dividirse en **geo** (tierra) + **gráfico** (dibujo/mapa). Por tanto, los datos geográficos contienen información de cualquier variable referenciada en un punto/área de la superficie terrestre y pueden representarse en mapas. El desarrollo de los datos geográficos ha producido grandes bases de datos espaciales y, a su vez, ha propiciado el desarrollo de herramientas para su tratamiento como los ya mencionados Sistemas de información geográficos y la Geocomputación.

¿Qué hace un Sistemas de información geográfico?

Un Sistema de información geográfica (SIG) es una herramienta que crea, administra, analiza y mapea todo tipo de datos. GIS conecta datos a un mapa, integrando datos de ubicación (**dónde** están las cosas) con todo tipo de información descriptiva (**cómo** son las cosas allí).

Esto proporciona una base para el mapeo y el análisis que se utiliza en la ciencia y en casi todas las industrias. GIS ayuda a los usuarios a comprender patrones, relaciones y contexto geográfico. Los beneficios incluyen una mejor comunicación y eficiencia, así como una mejor gestión y toma de decisiones.

La Fig. 2.1 muestra el flujo de trabajo de los SIG, que va desde (i) la elaboración de mapas, (ii) la obtención de geodatos o datos espaciales, (iii) el análisis de los datos geográficamente referenciados y (iv) la edición, mapeo y presentación de los resultados.

Pero es más, el desarrollo de la **Inteligencia Artificial** y la **Inteligencia computacional** se han convertido en herramientas creativas y complementarias a los convencionales GIS, dando origen a la *Geocomputación*, que trata de utilizar el *poder de los ordenadores para hacer cosas con los datos geográficos*.

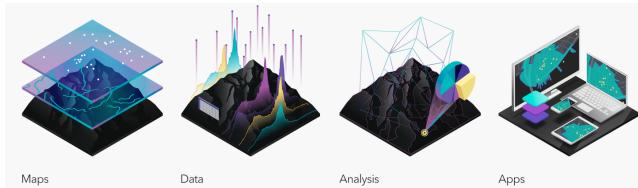


Figura 2.1: Flujo de trabajo de los GIS. Fuente: <https://www.esri.com/en-us/what-is-gis/overview>

¿Y que es la Geocomputación?

En primer lugar, señalar que, aunque la geocomputación es un término relativamente nuevo se encuentra influenciado por otros términos clásicos. De manera sencilla puede definirse como “*el proceso de aplicar tecnologías de computación a problemas geográficos*” [Rees, 1998]. Abrahart et al. [2000] aporta más elementos formales a esta definición destacando que “*la geocomputación trata sobre los diferentes tipos de geodatos, y sobre el desarrollo de geo-herramientas relevantes en un contexto científico*”.

La geocomputación está muy relacionada con otros términos como los SIG, ya definidos, y con diversos tipos de campos científicos, como las Geociencias, las Ciencias atmosféricas y climáticas, la Geoinformática, la Topología, la Ecología y las Ciencia de datos geográficos (GDS, Geographic Data Science).

Cada término comparte un énfasis en un enfoque **científico** (que implica reproducible y falsable) influenciado por los GIS, aunque sus orígenes y principales campos de aplicación difieren. La geocomputación es ampliamente utilizada en ámbitos como la sociología, el análisis político o el desarrollo de aplicaciones para móviles. Por tanto, usamos geocomputación como un sinónimo aproximado que encapsula a todas las ciencias que buscan usar datos geográficos para trabajos científicos aplicados.

¿Por que R para datos geográficos?

R es una herramienta con capacidades avanzadas de análisis, modelado y visualización. Por ejemplo, los nuevos entornos de desarrollo integrado (en inglés, Integrated Development Environment, **IDE**), como RStudio, han hecho que R sea más fácil de usar para muchos, facilitando la creación de mapas con un panel dedicado a la visualización interactiva [Lovelace et al., 2019]. Además, el uso del código R permite la enseñanza de la geocomputación con referencia a ejemplos reproducibles en lugar de conceptos abstractos. Por ejemplo, de una forma relativamente sencilla, se puede geoposicionar de manera interactiva la localización de la Puerta del Sol en Madrid y, además, dejar la el código R para hacerlo reproducible, ver Fig. 2.2.

```
library(leaflet)
leaflet(width = "100%", height = "500px") %>%
  addTiles() %>%
  setView(-3.703548, 40.417147, zoom = 16)
```



Figura 2.2: Localización interactiva de la Puerta del Sol en Madrid

Por otra parte R dispone de cientos de librerías especializadas para datos espaciales. Una descripción detallada puede ver se en [CRAN Task View: Analysis of Spatial Data](#)

Para no abrumar al lector, a continuación se muestran, de manera esquemática, las librerías más usadas para el tratamiento de datos espaciales y que se emplearán a lo largo de la asignatura Estadística Espacial y Espacio-Temporal, no sólo en el tema que nos ocupa:

- **sp** y **sf**: para el tratamiento de clases y métodos de los datos vectoriales.
- **raster,terra** y **stars** para datos raster.
- **gstat** y **geor**: para el análisis de datos geoestadísticos, ajuste y estimación de semivariogramas, interpolación, etc.

- `spdep` para el análisis de datos con modelos de econometría espacial, creación de matrices de contigüidad/distancia **W**, estimación de modelos econométricos espaciales, etc
- `spatstat` para el análisis de procesos de puntos espaciales, intensidad, etc.

2.2. Conceptos clave

Una vez visto el contexto actual de los datos georreferenciados y antes de entrar en detalle en su análisis, debemos tener en cuenta una serie de conceptos clave que se irán desarrollando a lo largo del tema.

Hemos dicho que Geográfico = Geo (tierra) + gráfico (mapa). Por tanto, si tenemos varios datos geográficos, localizados en distintos puntos de la tierra, es porque tenemos las **coordenadas** que los posicionan en esos puntos concretos. Asociado a estas coordenadas debemos conocer el **Sistema de referencia de espacial** o Coordinate reference system (CRS) en el que están proyectadas dichas coordenadas.

Por otra parte, los formatos de estos datos pueden ser **vectores** o **raster** como se explicará en la Sección 3.

Si damos un paso más e incorporamos el concepto de **distancia**, pues es lógico pensar que en un fenómeno de interés, por ejemplo, la modelización de la cantidad y dirección de lava en La Palma tras la erupción del volcán “Cumbre Vieja”, la distancia es un factor clave, pues aquellas zonas más cercanas al volcán tendrán niveles más parecidos entre sí y con valores más altos que aquellas que están más alejadas

En este caso el nivel de contaminación en el aire en La Palma no puede ser modelado como si las observaciones fuesen independientes pues las más cercanas entre sí serán más parecidas que las más lejanas, dando lugar al concepto de **dependencia espacial**. Y depende del tipo de datos espaciales tendremos tres grandes formas de abordar el tratamiento de los datos espaciales: **geoestadística**, **procesos de punto** y **econometría espacial** (véase sección 3.3).



Figura 2.3: Información espacial de la concentración de lava en Cumbre Vieja

Capítulo 3

Formatos de datos espaciales

3.1. Ficheros de datos espaciales

Los **datos espaciales**, también conocidos como datos **geoespaciales**, son aquellos datos relacionados o que contienen información de una localización o área geográfica de la superficie de la Tierra. En los últimos años se han desarrollado diversos estándares de ficheros que permiten almacenar y compatir este tipo de datos. En esta sección presentamos de una manera breve algunos de los tipos de archivos más habituales empleados en el almacenamiento de geodatos.

Información de vectores

Las características de los datos de vector se analizan en profundidad en la Sección 3.2.1. Las extensiones más habituales de los archivos que contienen datos de vectores son:

Cuadro 3.1: Ficheros con datos vector

Tipo	Extensión
Shapefile	.shp, .shx, .dbf
GeoPackage vector	.gPKG
GeoJson	.geojson
GPX	.gpx
Geography Markup Language	.gml
Keyhole Markup Language	.kml

Adicionalmente, es también común encontrarse con datos espaciales distribuidos en ficheros planos (.csv, .txt o incluso Excel). Generalmente este tipo de

Cuadro 3.2: Ejemplo: Coordenadas de Aeropuertos en España

Ciudad	longitud	latitud
Huesca	-0.323300	42.08080
Palma De Mallorca	2.702896	39.59862
Lillo / Toledo	-3.320560	39.71694
Valladolid	-4.850000	41.70111
Sabadell	2.105030	41.52090
Madrid	-3.785140	40.37067

datos corresponde a localizaciones (puntos) y la información geográfica suele especificarse en dos columnas que contienen las coordenadas de la localización.

Una nota importante sobre el uso de Shapefile

Shapefile surgió como uno de los primeros formatos de intercambio de datos geográficos a principios de los años 90 del siglo XX, y en la actualidad es quizá el formato más empleado. Sin embargo tiene una serie de limitaciones que es necesario conocer:

1. **Es un formato multiarchivo:** El formato shapefile se compone de diversos archivos que son necesarios para la correcta lectura de los datos. Es por ello que cuando se intercambia información han de proporcionarse como mínimo los archivos con extensiones **.shp**, **.shx**, **.dbf** (por ejemplo: **datos.shp**, **datos.shx**, **datos.dbf**). Lo mas habitual es distribuirlos en un fichero **.zip** del que se deben extraer todos estos archivos.
2. **El Sistema Geográfico de Coordenadas es opcional:** Como explicaremos en la Sección 3.3, esta información es crucial para trabajar con datos geográficos, sin embargo el estándar Shapefile lo considera opcional, siendo esta una limitación importante.

Además de estas limitaciones, existen otras relacionadas con el nombre y la longitud de los campos, o la capacidad de almacenamiento de los Shapefiles, que se pueden consultar en varios recursos web que profundizan en este tema (ver [mappingGIS](#), <http://switchfromshapefile.org/>). Un buen **reemplazo** de este formato es el formato **GeoPackage** (**.gpkg**), que no tiene este tipo de limitaciones.

Información de ráster

Las características de los datos ráster se analizan en profundidad en la Sección 3.2.2. Las extensiones más habituales de los archivos que contienen datos ráster son:

Cuadro 3.3: Ficheros con datos ráster

Tipo	Extensión
ASCII Grid	.asc
GeoTIFF	.tif, .tiff
Enhanced Compression Wavelet	.ecw

3.2. Modelo de datos

En el ámbito del análisis espacial en R, se pueden clasificar el formato de datos espaciales en función del modelo de datos [Lovelace et al., 2019]. Se pueden distinguir dos tipos de modelos de datos: **vectores** y **ráster**.

3.2.1. Datos de vectores

Este modelo está basado en puntos georeferenciados. Los **puntos** pueden representar localizaciones específicas, como la localización de edificios:

```
library(ggplot2)
library(sf)

# Hospitales en Toledo segun Eurostat
hosp_toledo <- st_read("data/hosp_toledo.geojson", quiet = TRUE)

# Plot
ggplot() +
  geom_sf(
    data = hosp_toledo, aes(fill = "Centros Sanitarios"),
    color = "blue"
  ) +
  labs(
    caption = "Datos: Eurostat",
    title = "Hospitales y Centros de Salud en Toledo",
    fill = ""
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")
```

Estos puntos también pueden estar conectados entre sí, de manera que formen geometrías más complejas, como **líneas** y **polígonos**:

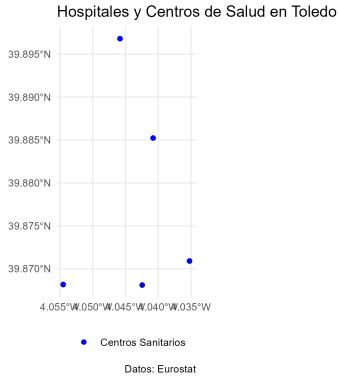


Figura 3.1: Datos vector: Puntos

```
tajo <- st_read("data/tajo_toledo.shp", quiet = TRUE)
toledo <- st_read("data/toledo_ciudad.gpkg", quiet = TRUE)

ggplot(toledo) +
  geom_sf(fill = "cornsilk2") +
  geom_sf(data = tajo, col = "lightblue2", lwd = 2, alpha = 0.7)
  +
  geom_sf(data = hosp_toledo, col = "blue") +
  coord_sf(
    xlim = c(-4.2, -3.8),
    ylim = c(39.8, 39.95)
  ) +
  theme_minimal()
```

En la Fig. 3.2, el río Tajo está representado como una línea (sucesión de puntos unidos entre sí) y la ciudad de Toledo como un polígono (línea de puntos cerrada formando un continuo). A modo ilustrativo, la Fig. 3.3 representa la descomposición en puntos de todos los datos espaciales representados en la Fig. 3.2.

3.2.2. Datos ráster

Los datos ráster son datos representados en una rejilla rectangular de píxeles (denominada **matriz**) que se puede visualizar en diversos dispositivo de representación. El caso más cotidiano de un ráster es una fotografía, donde la imagen se representa como una serie de celdas, determinadas por la resolución

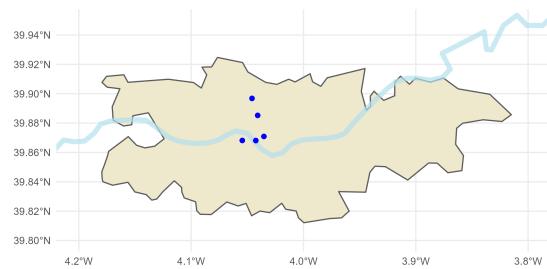


Figura 3.2: Datos vector: Puntos, líneas y polígonos

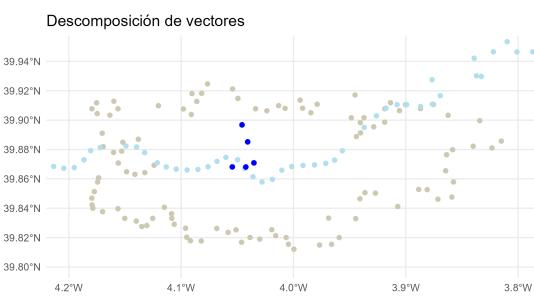


Figura 3.3: Datos vector: Descomposición en puntos

de la imagen (número total de píxeles, determinados como número de píxeles en cada fila por número de píxeles en cada columna) y el color que presenta cada uno de estos píxeles.

En el ámbito de los datos espaciales, la definición es muy similar. Un archivo ráster está formado por una malla regular de píxeles georreferenciada, tal y como muestra la Fig. 3.4:

```
library(raster)

elev <- raster("data/Toledo_DEM.tif")
plot(elev, main = "Elevación de la provincia de Toledo")

# Mostramos el grid
polys <- rasterToPolygons(elev)
plot(polys, add = TRUE, border = "grey90")

# Añadimos la provincia
Tol_prov <- st_read("data/Toledo_prov.gpkg", quiet = TRUE)

# Si queremos solamente la forma en sf, usamos st_geometry
plot(st_geometry(Tol_prov), add = TRUE)
```

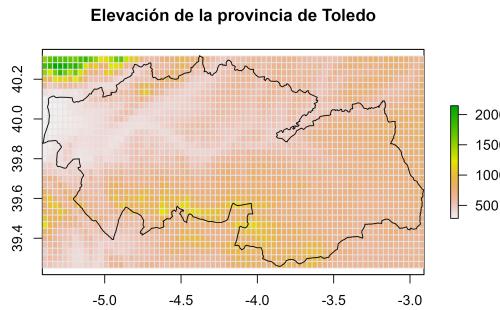


Figura 3.4: Datos ráster

En la Fig. 3.4, el objeto ráster `elev` tiene únicamente una capa (denominada `ESP_alt`). Eso implica que cada píxel tiene asociado un único valor, en este caso, en este caso la altitud media del terreno observada:

Los rásters pueden contener varias capas (o layers), de manera que cada píxel puede tener asociados varios valores. Volviendo al ejemplo de la fotografía, en un modelo simple de color RGB cada píxel lleva asociado 3 valores (rojo, verde

Cuadro 3.4: Datos de un ráster (detalle)

x	y	Toledo DEM
-5.391667	40.3	1498.312
-5.358333	40.3	1701.125
-5.325000	40.3	1825.312
-5.291667	40.3	1739.062
-5.258333	40.3	1756.062
-5.225000	40.3	1659.688
-5.191667	40.3	1607.375
-5.158333	40.3	1809.562
-5.125000	40.3	1874.625
-5.091667	40.3	1691.312
-5.058333	40.3	1511.500
-5.025000	40.3	1207.000
-4.991667	40.3	1160.125
-4.958333	40.3	1396.125
-4.925000	40.3	1624.125

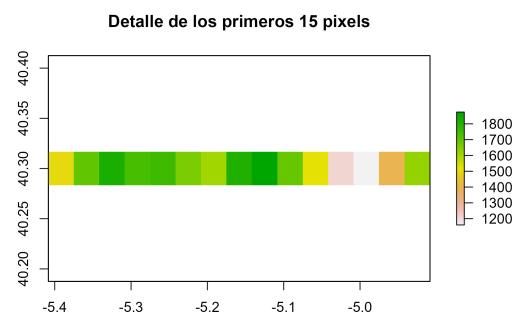


Figura 3.5: Datos ráster: Detalle

Cuadro 3.5: Datos de un ráster multicapa (detalle)

x	y	lyr.1	lyr.2	lyr.3
-5.466412	40.34418	215.2128	208.1061	190.5410
-5.463875	40.34418	228.0369	223.1854	211.2115
-5.461338	40.34418	229.3495	224.3414	213.4325
-5.458800	40.34418	215.8592	208.8660	191.2922
-5.456263	40.34418	219.2696	212.8231	196.6812
-5.453725	40.34418	235.0954	231.4222	222.4115
-5.451188	40.34418	240.3514	237.9094	231.4736
-5.448651	40.34418	237.2358	233.7561	226.2005
-5.446113	40.34418	229.9570	225.3262	214.6201
-5.443576	40.34418	226.7812	221.6796	209.2929
-5.441038	40.34418	222.3593	216.5022	202.0188
-5.438501	40.34418	220.9312	214.9060	200.0306
-5.435964	40.34418	224.7755	219.2661	206.2156
-5.433426	40.34418	222.0479	216.0124	201.6103
-5.430889	40.34418	225.0516	219.8074	207.0263

o azul), de manera que al combinar las tres capas se puede definir un color distinto en cada píxel.

En la Fig. 3.6 vamos a usar una imagen de mapa georreferenciada, como las proporcionadas por servicios de mapas online, para analizar su composición.

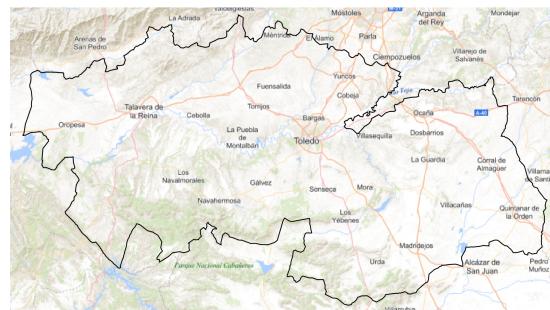


Figura 3.6: Datos ráster con varias bandas

El ráster se puede descomponer en las tres capas RGB mencionadas anteriormente:



Figura 3.7: Datos ráster multicapa: Descomposición

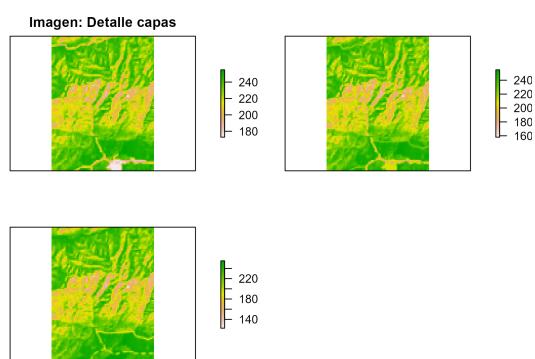


Figura 3.8: Datos ráster multicapa: Descomposición

3.3. Sistema de Referencia de Coordenadas (CRS)

Un sistema de referencia de coordenadas (o CRS por sus siglas en inglés, **Coordinate Reference System**) permite relacionar datos espaciales con su localización en la superficie terrestre.

Los CRS constituyen por tanto un aspecto fundamental en el análisis y representación de datos espaciales, ya que nos permiten identificar con exactitud la posición de los datos sobre el globo terráqueo.

Así mismo, cuando se trabaja con datos espaciales provenientes de distintas fuentes de información, es necesario comprobar que dichos datos se encuentran definidos en el mismo CRS:

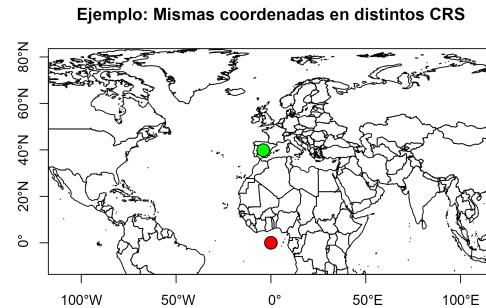


Figura 3.9: Representación de mismos valores de coordenadas en distintos CRS

En la Fig. 3.9, ambos puntos (verde y rojo) tienen los mismos valores de coordenadas en los ejes X e Y, en este caso las correspondientes a la ciudad de Toledo. Sin embargo, presentan distintos CRS. Por este motivo, al representar ambos puntos en un mapa, se observa que no se están refiriendo a la misma localización geográfica. Esto es así porque el CRS define la referencia (punto $x=0$ e $y = 0$) y las unidades de los ejes (grados, metros, millas).

Como conclusión, **además de disponer de las coordenadas de los datos espaciales, es necesario conocer el CRS en el que están definidos para conocer de manera exacta su localización geográfica**. Además, nótese que para cualquier **análisis de datos espaciales** es necesario que todos los geodatos **se encuentren referenciados en el mismo CRS**. Esto se consigue transformando (o proyectando) los datos a un CRS común, nunca sobreescribiendo el CRS de los mismos.

3.3.1. Tipos de CRS

A continuación se definen los dos grandes tipos de CRS, los CRS geográficos y los CRS proyectados.

3.3.1.1. CRS geográficos

Los CRS geográficos son aquellos en los que los parámetros empleados para localizar una posición espacial son la latitud y la longitud:

- **Latitud:** Es la distancia angular expresada en grados sobre el plano definido por el ecuador terrestre. Determina la posición sobre de una localización en el eje Norte-Sur de la Tierra y toma valores en el rango $[-90, 90]$. Las líneas imaginarias determinadas por una sucesión de puntos con la misma latitud a lo largo del eje Este-Oeste se denominan **paralelos** (Ver Fig. 3.10).
- **Longitud:** Es la distancia angular expresada en grados sobre el plano definido por el meridiano de Greenwich. Determina la posición sobre de una localización en el eje Este-Oeste de la Tierra y toma valores en el rango $[-180, 180]$. Las líneas imaginarias determinadas por una sucesión de puntos con la misma longitud a lo largo del eje Este-Oeste se denominan **meridianos** (Ver Fig. 3.10).

Es muy importante destacar que en un sistema de coordenadas geográfico, es decir, basado en latitudes y longitudes, las **distancias** entre dos puntos representan **distancias angulares**. Por ejemplo, la distancia entre el meridiano de Greenwich y el meridiano correspondiente a la longitud 20° siempre es de $+20^\circ$. Sin embargo, debido a la forma esférica de la Tierra, la longitud en metros entre ambos meridianos no es constante.

3.3.1.2. CRS proyectados

La representación de formas tridimensionales en un soporte plano (dos dimensiones) presenta algunos retos. Por ello, es habitual trabajar con proyecciones de mapas.

Una **proyección geográfica** es un método para reducir la superficie de la esfera terrestre a un sistema cartesiano de dos dimensiones. Para ello, es necesario transformar las coordenadas longitud y latitud en coordenadas cartesianas x e y.

Es importante destacar que las proyecciones pueden incluir un punto de origen ($X=0, Y=0$) y unas unidades de distancia (habitualmente metros) específicas. Por ejemplo, la **proyección cónica equiáreas de Albers** (específica para

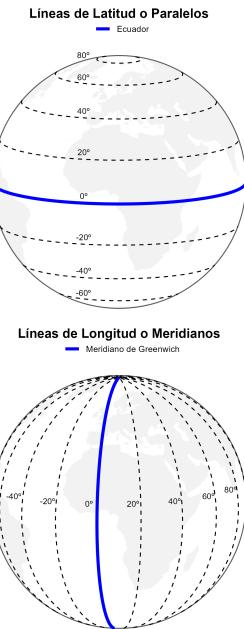


Figura 3.10: Paralelos y Meridianos terrestres

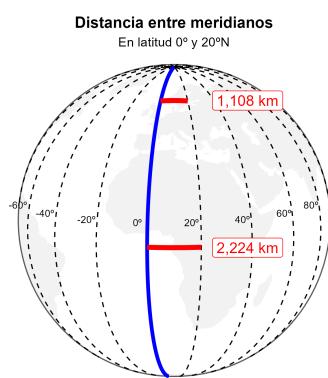


Figura 3.11: Distancia entre meridianos en distintas latitudes

Estados Unidos) define su punto de referencia (0,0) en la latitud 40° N y longitud 96°, y la unidad de variación están definida en metros. De ahí la importancia de conocer el CRS de los datos geográficos, como se expuso al principio de este tema.

El Anexo A proporciona más información sobre los tipos de CRS proyectados.

3.3.2. Trabajando con proyecciones en R

Existe toda una serie de proyecciones predefinidas, identificadas mediante los **códigos EPSG, ESRI, WKT** o proj4 (en desuso en R, pero todavía admitidos). Existen varios recursos web donde se pueden consultar y seleccionar los códigos correspondientes:

- <https://epsg.io/>
- <https://spatialreference.org/>
- <https://proj.org/operations/projections/index.html>

Algunos de los códigos de proyecciones que es fundamental conocer son:

- **EPSG: 4326:** Proyección correspondiente a WGS 84, que es el sistema usado por los sistemas GPS. Cuando trabajemos con coordenadas geográficas longitud/latitud, este es habitualmente el CRS de referencia.
- **EPSG: 3857:** Código correspondiente a la proyección de Mercator, usada habitualmente por servicios como Google Maps, etc.

Se pueden consultar otros CRS de uso común en España en la página del [Ministerio de Agricultura, Pesca y Alimentación](#)

En la sección 3.3.3 veremos cómo encontrar un CRS usando el paquete `crsuggest`.

El paquete `sf` permite obtener los parámetros de cualquier proyección mediante la función `st_crs()`:

(i) EPSG WGS 84 (Sistema Global GPS): EPSG 4326

```
library(sf)

# Ejemplo: EPSG WGS 84 (Sistema Global GPS): EPSG 4326
st_crs(4326)
#> Coordinate Reference System:
#>   User input: EPSG:4326
```

```

#>   wkt:
#> GEOGCRS["WGS 84",
#>   DATUM["World Geodetic System 1984",
#>     ELLIPSOID["WGS 84",6378137,298.257223563,
#>     LENGTHUNIT["metre",1]],
#>   PRIMEM["Greenwich",0,
#>     ANGLEUNIT["degree",0.0174532925199433]],
#>   CS[ellipsoidal,2],
#>     AXIS["geodetic latitude (Lat)",north,
#>       ORDER[1],
#>       ANGLEUNIT["degree",0.0174532925199433]],
#>     AXIS["geodetic longitude (Lon)",east,
#>       ORDER[2],
#>       ANGLEUNIT["degree",0.0174532925199433]],
#>   USAGE[
#>     SCOPE["Horizontal component of 3D system."],
#>     AREA["World."],
#>     BBOX[-90,-180,90,180]],
#>   ID["EPSG",4326]

```

(ii) ESRI North America Albers Equal Area Conic: ESRI:102008

```

# Usando código ESRI North America Albers Equal Area Conic

st_crs("ESRI:102008")
#> Coordinate Reference System:
#>   User input: ESRI:102008
#>   wkt:
#> PROJCRS["North_America_Albers_Equal_Area_Conic",
#>   BASEGEOGCRS["NAD83",
#>     DATUM["North American Datum 1983",
#>       ELLIPSOID["GRS 1980",6378137,298.257222101,
#>       LENGTHUNIT["metre",1]],
#>     PRIMEM["Greenwich",0,
#>       ANGLEUNIT["Degree",0.0174532925199433]]],
#>   CONVERSION["North_America_Albers_Equal_Area_Conic",
#>     METHOD["Albers Equal Area",
#>       ID["EPSG",9822]],
#>     PARAMETER["Latitude of false origin",40,
#>       ANGLEUNIT["Degree",0.0174532925199433],
#>       ID["EPSG",8821]],
#>     PARAMETER["Longitude of false origin",-96,
#>       ANGLEUNIT["Degree",0.0174532925199433],
#>       ID["EPSG",8822]],
#>     PARAMETER["Latitude of 1st standard parallel",20,

```

```

#>           ANGLEUNIT["Degree",0.0174532925199433],
#>           ID["EPSG",8823]],
#>           PARAMETER["Latitude of 2nd standard parallel",60,
#>           ANGLEUNIT["Degree",0.0174532925199433],
#>           ID["EPSG",8824]],
#>           PARAMETER["Easting at false origin",0,
#>           LENGTHUNIT["metre",1],
#>           ID["EPSG",8826]],
#>           PARAMETER["Northing at false origin",0,
#>           LENGTHUNIT["metre",1],
#>           ID["EPSG",8827]],
#>           CS[Cartesian,2],
#>           AXIS["(E)",east,
#>           ORDER[1],
#>           LENGTHUNIT["metre",1]],
#>           AXIS["(N)",north,
#>           ORDER[2],
#>           LENGTHUNIT["metre",1]],
#>           USAGE[
#>               SCOPE["Not known."],
#>               AREA["North America - onshore and offshore: Canada -
#>           Alberta; British Columbia; Manitoba; New Brunswick;
#>           Newfoundland and Labrador; Northwest Territories; Nova
#>           Scotia; Nunavut; Ontario; Prince Edward Island; Quebec;
#>           Saskatchewan; Yukon. United States (USA) - Alabama; Alaska
#>           (mainland); Arizona; Arkansas; California; Colorado;
#>           Connecticut; Delaware; Florida; Georgia; Idaho; Illinois;
#>           Indiana; Iowa; Kansas; Kentucky; Louisiana; Maine; Maryland;
#>           Massachusetts; Michigan; Minnesota; Mississippi; Missouri;
#>           Montana; Nebraska; Nevada; New Hampshire; New Jersey; New
#>           Mexico; New York; North Carolina; North Dakota; Ohio;
#>           Oklahoma; Oregon; Pennsylvania; Rhode Island; South Carolina;
#>           South Dakota; Tennessee; Texas; Utah; Vermont; Virginia;
#>           Washington; West Virginia; Wisconsin; Wyoming."],
#>               BBOX[23.81,-172.54,86.46,-47.74]],
#>               ID["ESRI",102008]]

```

(iii) Usando proj4string: Robinson: +proj=robin

```

# Usando proj4string: Robinson

st_crs("+proj=robin")
#> Coordinate Reference System:
#>   User input: +proj=robin
#>   wkt:

```

```

#> PROJCRS["unknown",
#>     BASEGEOGCRS["unknown",
#>         DATUM["World Geodetic System 1984",
#>             ELLIPSOID["WGS 84",6378137,298.257223563,
#>                 LENGTHUNIT["metre",1]],
#>             ID["EPSG",6326]],
#>         PRIMEM["Greenwich",0,
#>             ANGLEUNIT["degree",0.0174532925199433],
#>             ID["EPSG",8901]],
#>         CONVERSION["unknown",
#>             METHOD["Robinson"],
#>             PARAMETER["Longitude of natural origin",0,
#>                 ANGLEUNIT["degree",0.0174532925199433],
#>                 ID["EPSG",8802]],
#>             PARAMETER["False easting",0,
#>                 LENGTHUNIT["metre",1],
#>                 ID["EPSG",8806]],
#>             PARAMETER["False northing",0,
#>                 LENGTHUNIT["metre",1],
#>                 ID["EPSG",8807]]],
#>         CS[Cartesian,2],
#>             AXIS["(E)",east,
#>                 ORDER[1],
#>                 LENGTHUNIT["metre",1,
#>                     ID["EPSG",9001]]],
#>             AXIS["(N)",north,
#>                 ORDER[2],
#>                 LENGTHUNIT["metre",1,
#>                     ID["EPSG",9001]]]

```

La mayoría de los objetos espaciales serán de la clase `sf`, por tanto, resulta interesante conocer cómo se proyectan estos objetos.

Es posible proyectar un objeto `sf` mediante la función `st_transform()`. En el siguiente ejemplo vemos cómo partimos de un objeto con **EPSG:4326** y cambiamos su proyección a otras proyecciones, como **Mercator** o **Robinson**:

```

# Usa datos del paquete giscoR

library(giscoR)

paises <- gisco_get_countries()

# Comprobamos el CRS de estos datos

```

```

# Se puede almacenar en un objeto y usar posteriormente
# Vemos que es EPSG:4326, por tanto son coordenadas geográficas
#<-- longitud/latitud
st_crs(paises)
#> Coordinate Reference System:
#>   User input: EPSG:4326
#>   wkt:
#>     GEOGCS["WGS 84",
#>       DATUM["WGS_1984",
#>         SPHEROID["WGS 84",6378137,298.257223563,
#>           AUTHORITY["EPSG", "7030"]],
#>         AUTHORITY["EPSG", "6326"]],
#>       PRIMEM["Greenwich",0,
#>         AUTHORITY["EPSG", "8901"]],
#>       UNIT["degree",0.0174532925199433,
#>         AUTHORITY["EPSG", "9122"]],
#>       AUTHORITY["EPSG", "4326"]]

# Plot
plot(st_geometry(paises), axes = TRUE)

```

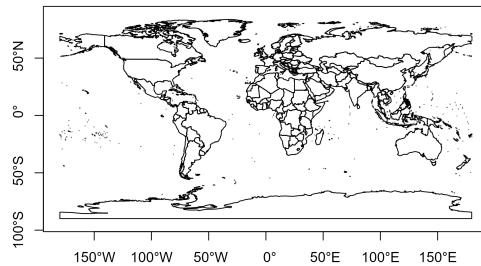


Figura 3.12: Proyección del mundo en coordenadas geográficas (EPSG 4326)

```

# Proyectamos a Mercator
# El eje cambia porque Mercator usa metros
paises_merc <- st_transform(paises, st_crs(3857))
plot(st_geometry(paises_merc), axes = TRUE)

```

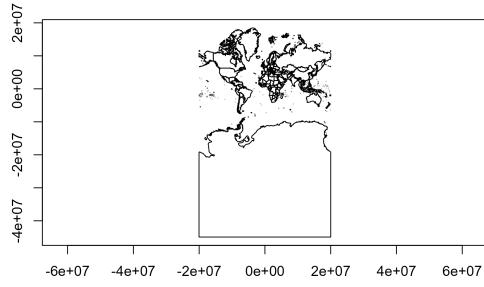


Figura 3.13: Proyección del mundo en Mercator (EPSG 3857)

```
# Proyectamos a Robinson
paises_robin <- st_transform(paises, st_crs("+proj=robin"))
plot(st_geometry(paises_robin), axes = TRUE)
```

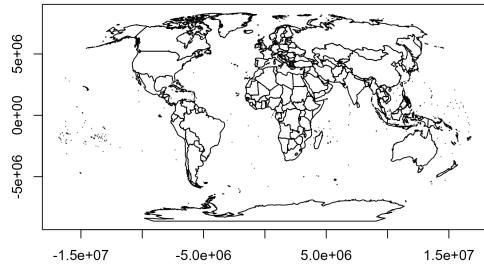


Figura 3.14: Proyección del mundo en Robinson (+proj=robin)

Como se comentó anteriormente, cuando se usan geodatos de diversas fuentes, es necesario que todos presenten el mismo CRS. En la Fig 3.15 se muestra lo que ocurre si esto no se cumple:

```
# Añadimos a este mapa puertos mundiales de giscoR
puertos <- gisco_get_ports()
plot(st_geometry(paises_robin), main = "Puertos en el mundo")
plot(st_geometry(puertos), add = TRUE, col = "red", pch = 20)
```



Figura 3.15: Ejemplo: Puertos del mundo

Vemos que ha habido algún tipo de error, ¿a que puede deberse?

```
# Comprueba CRS
st_crs(puertos) == st_crs(paises_robin)
#> [1] FALSE

# Los puertos no están en Robinson! Proyectamos al mismo CRS
puertos_robin <- st_transform(puertos, st_crs(paises_robin))
plot(st_geometry(paises_robin), main = "Puertos en el mundo")
plot(st_geometry(puertos_robin), add = TRUE, col = "blue", pch =
  20)
```



Figura 3.16: Ejemplo: Puertos del mundo, CRS alineados

Como vemos, en el primer mapa (Fig. 3.15) los puertos se concentran en un único

punto, dado que no están referenciados en el mismo CRS. Tras proyectarlos al mismo CRS, el mapa se representa adecuadamente (Fig. 3.16).

En otros paquetes, como **sp** o **raster**, existen funciones parecidas que nos van a permitir obtener los parámetros de un CRS y proyectar los objetos al CRS deseado. Cuando empleemos el paquete **sp** podemos usar las funciones **CRS()** y **spTransform()**:

```
library(sp)

# Convertimos sf a sp
paises_sp <- as(paises, "Spatial")

# En sp podemos usar:
# CRS("+proj=robin")
#
# O también desde sf
# CRS(st_crs(paises_robin)$proj4string)

paises_sp_robin <- spTransform(paises_sp, CRS("+proj=robin"))
plot(paises_sp_robin)
```

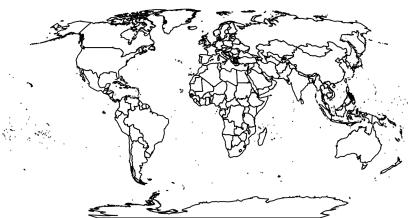


Figura 3.17: Transformaciones en **sp**

En el caso de un objeto **raster**, podemos usar **crs()** y **projectRaster()**:

```
library(raster)

# Extrae información de altitud para España
```

```

elev <- raster("data/ESP_msk_alt.grd")

# Transforma
elev_robinson <- projectRaster(elev, crs = crs("+proj=robin"))
plot(elev_robinson)

```

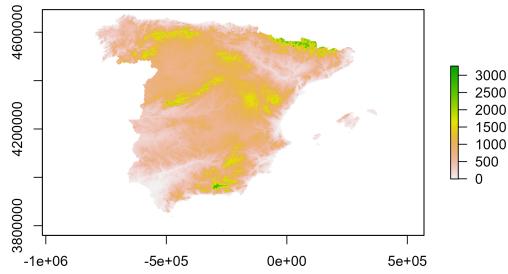


Figura 3.18: Transformaciones en raster

Por último, en el paquete `terra` las funciones correspondientes son `crs()` y `project()`:

```

library(terra)

# Convierte de raster a terra
elev_terra <- rast(elev)

# Transforma
elev_terra_robinson <- terra::project(elev_terra,
                                         crs("+proj=robin"))
plot(elev_terra_robinson)

```

3.3.3. ¿Qué proyección uso?

El CRS adecuado para cada análisis depende de la localización y el rango espacial de los datos. Un CRS adecuado para representar un mapa del mundo puede no serlo para representar datos de zonas específicas de la Tierra. Los recursos

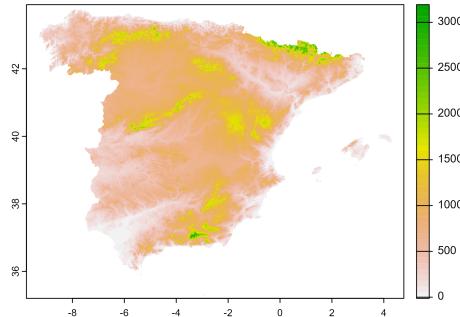


Figura 3.19: Transformaciones en terra

Cuadro 3.6: Tabla sugerencias, detalle

crs_code	crs_name	crs_type	crs_gcs	crs_units	crs_proj4
2062	Madrid 1870 (Madrid) / Spain LCC	projected	4903	m	+proj=lcc +lat_1=40 +lat_0=38 +lon_0=-3.7 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs
2154	RGF93 / Lambert-93	projected	4171	m	+proj=lcc +lat_0=46 +lon_0=-2.33333 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs
26191	Merchich / Nord Maroc	projected	4261	m	+proj=lcc +lat_1=33 +lat_0=30 +lon_0=-4 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs
3944	RGF93 / CC44	projected	4171	m	+proj=lcc +lat_0=46 +lon_0=-2.33333 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs
3943	RGF93 / CC43	projected	4171	m	+proj=lcc +lat_0=46 +lon_0=-2.33333 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs
27573	NTF (Paris) / Lambert zone III	projected	4807	m	+proj=lcc +lat_1=46 +lat_0=40 +lon_0=-2.33333 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs
27572	NTF (Paris) / Lambert zone II	projected	4807	m	+proj=lcc +lat_1=40 +lat_0=36 +lon_0=-2.33333 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs
27563	NTF (Paris) / Lambert Sud France	projected	4807	m	+proj=lcc +lat_1=36 +lat_0=30 +lon_0=-2.33333 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs
30791	Nord Sahara 1959 / Nord Algerie	projected	4307	m	+proj=lcc +lat_1=36 +lat_0=30 +lon_0=-2.33333 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs
30493	Voirol 1879 / Nord Algerie (ancienne)	projected	4671	m	+proj=lcc +lat_1=36 +lat_0=30 +lon_0=-2.33333 +x_0=600000 +y_0=600000 +ellps=GRS80 +towgs84=420,-120,250 +units=m +no_defs

web mencionados anteriormente permiten la búsqueda de CRS por zona geográfica, y adicionalmente en **R** existe el paquete **crssuggest** [Walker, 2021] que nos facilita la labor, sugiriendo el CRS más adecuado para cada zona:

```
library(crsuggest)

# Usando raster
sugerencias <- suggest_crs(elev)

# Probamos sugerencia
crs_suggest <- suggest_crs(elev, limit = 1)

elev_suggest <- projectRaster(elev, crs =
  ↪ raster::crs(crs_suggest$crs_proj4))
```

```
plot(elev_suggest)
```

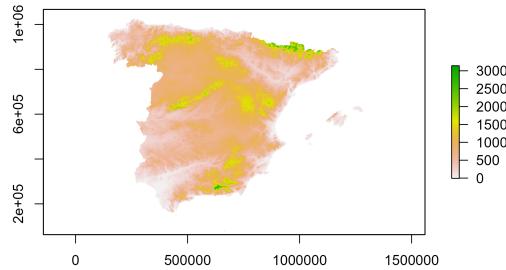


Figura 3.20: raster: Ejemplo de transformación usando crs_suggest

```
# Ejemplo con sf: China

china <- gisco_get_countries(country = "China")
china_crs <- suggest_crs(china, limit = 1)

china_crs
#> # A tibble: 1 x 6
#>   crs_code crs_name           crs_type
#>   <chr>    <chr>            <chr>
#>   ↘ crs_gcs crs_units crs_proj4
#>   <chr>    <chr>            <chr>
#>   ↘ <dbl>    <chr>            <chr>
#>   1 4584     New Beijing / Gauss-Kruger CM 105E projected
#>   ↘ 4555 m      +proj=tmerc +lat_0=0 +lo~

china_suggest <- st_transform(
  china,
  st_crs(as.integer(china_crs$crs_code))
)

plot(st_geometry(china_suggest), axes = TRUE)
```

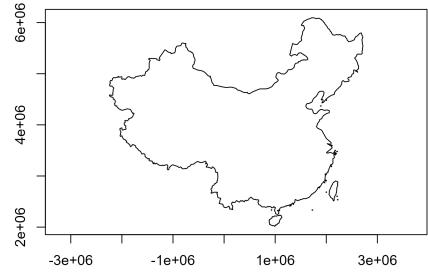


Figura 3.21: sf: Ejemplo de transformación usando crssuggest

Capítulo 4

Estadística espacial

La estadística espacial reconoce y aprovecha la ubicación espacial de los datos a la hora de diseñar, recopilar, gestionar, analizar y mostrar las observaciones. Éstas son generalmente **dependientes**, si bien existen modelos espaciales a disposición del investigador que permiten tratar con dicha dependencia espacial a la hora de llevar a cabo labores de predicción. Por extensión, la estadística espacio-temporal incorpora, además, el tiempo y su interacción con el espacio como argumento de ayuda en tales labores predictivas.

Las mediciones y modelos espaciales están presentes, sorprendentemente, en una amplia variedad de disciplinas científicas. Los orígenes de la vida humana vinculan los estudios de la evolución de las galaxias, la estructura de las células biológicas y los patrones de asentamiento arqueológico. Los ecólogos estudian las interacciones entre plantas y animales. Silvicultores y agricultores necesitan investigar las variaciones que se producen en el terreno para sus experimentos. La estimación de las precipitaciones y de las reservas de oro y petróleo es de vital importancia económica. Estos son, entre otros, buenos ejemplos de la importancia del espacio (espacio-tiempo en su caso) en el mundo de la Ciencia.

Sin embargo, el estudio de la **variabilidad espacial**, y sobre todo espacio-temporal, es una disciplina relativamente nueva en el marco de la Estadística, lo que explica la escasez de instrumentos de estadística espacial 30 años atrás. En los últimos 10 años ha habido una creciente toma de conciencia de esta necesidad, habiéndose realizado un gran esfuerzo por buscar herramientas adecuadas y útiles a tales efectos. Y todo ello porque utilizar modelos espaciales o espacio-temporales para caracterizar y explotar la dependencia espacial (o espacio-temporal) de un conjunto de observaciones tiene importantes ventajas ([Montero et al. \[2011\]](#)):

1. Modelos más generales, ya que, en la mayoría de los casos, los modelos clásicos que no tienen en consideración la dimensión espacial o la interac-

ción de las dimensiones espacial y temporal son un caso particular de un modelo espacial o espacio-temporal.

2. Estimaciones más eficientes: de la tendencia, de los efectos de las variables explicativas, de promedios regionales,...
3. Mejora de las predicciones: más eficientes, con propiedades de extrapolación más estables,...
4. La variación espacial no explicada en la estructura de la media debe ser absorbida por la estructura del error, por lo que un modelo que incorpore la dependencia espacial puede decirse que está protegido frente a una mala especificación de este tipo. Esto, en muchos casos, tiene como resultado una simplificación en la especificación de la tendencia; en general, los modelos con dependencia espacial suelen tener una descripción más parsimoniosa (en ocasiones con muchos menos parámetros) que los clásicos modelos de superficie de tendencia.

4.1. Antes de continuar... dependencia espacial.

Frecuentemente los datos tienen una componente espacial y/o temporal asociada a ellos y es de esperar que datos cercanos en el espacio o en el tiempo sean más semejantes que aquellos que están más alejados; en cuyo caso **no** deben ser modelados como estadísticamente independiente, sino que habrá que tomar en cuenta esa dependencia espacial o espacio-temporal.

De forma natural y de acuerdo a la Ley Tobler (1973) surge la idea de que los datos cercanos en el espacio o en el tiempo serán más similares y estarán más correlacionados entre sí que aquellos que están más lejanos. Además, esta correlación disminuye al aumentar la separación entre ellos, por lo que se puede pensar en la presencia de una dependencia espacial o espacio-temporal. Esto da lugar al concepto de proceso espacial o espacio-temporal.

Si los datos no exhiben dependencia espacial no tiene sentido aplicar las herramientas de estadística espacial. Veamos un ejemplo simulado de unos datos que muestran dependencia espacial y otros puramente aleatorios.

La Fig. 4.1 muestra unos datos simulados que presentan una estructura de dependencia espacial (panel izquierdo) frente a unos datos totalmente aleatorios (panel derecho), en ambos casos distribuidos de forma irregular en el espacio.

La Fig. 4.2, al igual que la Fig. 4.1 presenta unos datos simulados donde que presentan una estructura de dependencia espacial (panel izquierdo) frente a unos datos totalmente aleatorios (panel derecho), pero en este caso estos datos se distribuyen de forma ordenada en el espacio a través de una rejilla regular.

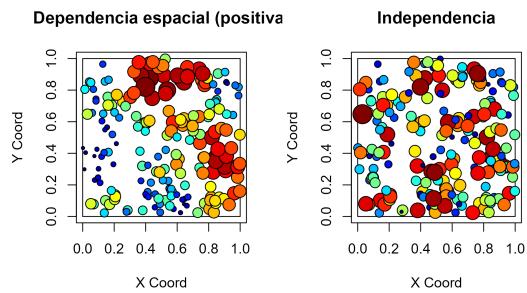


Figura 4.1: Puntos: Ejemplo de independencia espacial

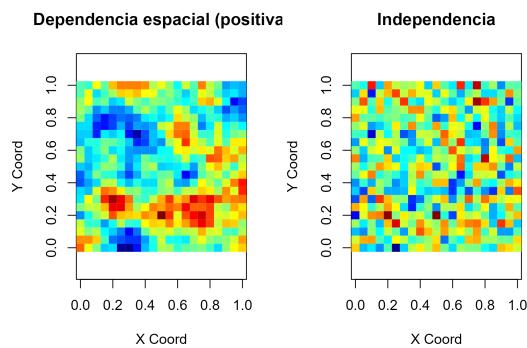


Figura 4.2: Rejilla: Ejemplo de independencia especial

4.2. Datos espaciales

Los **datos espaciales**, también conocidos como datos **geoespaciales**, son aquellos datos relacionados o que contienen información de una localización o área geográfica de la superficie de la Tierra.

La forma más intuitiva de representar los datos espaciales es a través de un mapa.

```
# Mapa de porcentaje de mujeres en Castilla-La Mancha

library(mapSpain)

# Datos de población
pob <- mapSpain::pobmun19

# Datos en forma de tabla, sin información en formato espacial
# head(pob)

# Porcentaje
pob$porc_mujeres <- pob$women / pob$pob19 * 100

# Datos espaciales
geo <- esp_get_munic(region = "Castilla-La Mancha")

# Estos datos tienen una columna (geometry) con coordenadas.
# head(geo)

# Une ambos datos
geo_pob <- merge(geo,
  pob,
  by = c("cpro", "cmun"),
  all.x = TRUE
)

# Mapa básico
plot(geo_pob["porc_mujeres"],
  # Cambiamos título
  main = "Castilla-La Mancha: % mujeres (2019)",

  # Cambiamos la paleta de colores para hacerlo mas atractivo
  # border = NA,
  pal = hcl.colors(12, "RdYlBu")
)
```

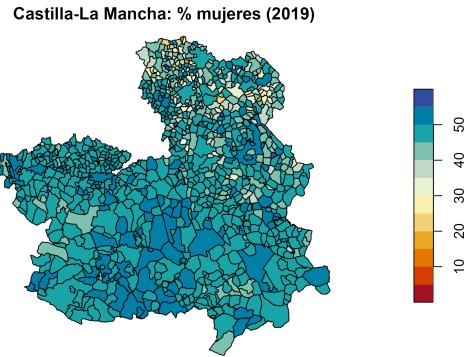


Figura 4.3: Porcentaje de Mujeres en Castilla-La Mancha

La Fig. 4.3 presenta una serie de elementos gráficos, característicos de los objetos espaciales:

- Los municipios de Castilla-la Mancha están representados por polígonos con un contorno negro y se rellenan de colores de acuerdo con la variable que estamos analizando, el porcentaje de mujeres en los municipios de Castilla-La Mancha en el año 2019.
- Una leyenda explica el significado de los colores.
- La variable, el porcentaje de mujeres en los municipios de Castilla-La Mancha en el año 2019, no parece distribuirse de manera independiente sino todo lo contrario, muestra un patrón espacial. Los municipios de las provincias Guadalajara y Cuenca (noreste) presentan tasas más bajas que los municipios del centro de la Comunidad.

```
head(geo_pob[["porc_mujeres"]])
#> Simple feature collection with 6 features and 1 field
#> Geometry type: POLYGON
#> Dimension: XY
#> Bounding box: xmin: -2.18037 ymin: 38.5441 xmax: -1.31112
#> ymax: 39.35597
#> Geodetic CRS: ETRS89
#> porc_mujeres           geometry
#> 1      52.02532 POLYGON ((-1.58316 39.20446...
#> 2      43.93064 POLYGON ((-1.40607 39.12384...
#> 3      51.14089 POLYGON ((-2.0562 38.88697, ...
#> 4      48.55491 POLYGON ((-1.54055 38.61066...
#> 5      48.78419 POLYGON ((-1.38514 39.35429...
#> 6      44.49541 POLYGON ((-2.15635 38.71074...
```

Antes de dibujar la Fig. 4.3 tuvimos que leer los datos de la librería `mapSpain` [Hernández, 2022] que contenía tanto la variable que hemos analizado como el formato del mapa. Tras unir variable y mapa con la función `merge()`. Al llevar a cabo un resumen del objeto espacial nos encontramos con las siguiente información:

- el conjunto de datos (seleccionado) tiene 919 registros (municipios)
- el tipo de geometría es POLYGON.
- el CRS es ETRS89

4.3. Clasificación de datos espaciales

Tal y como acabamos de señalar y de acuerdo con Schabenberger y Gotway (2005, p. 6), debido a que los datos espaciales surgen en una gran variedad de campos y aplicaciones, también hay una gran variedad de tipos de datos espaciales, estructuras y escenarios. Por tanto, una clasificación exhaustiva de los datos espaciales sería un reto muy difícil y hemos apostado por una clasificación general, simple y útil de datos espaciales proporcionada por Cressie [1993].

La **clasificación** de Cressie de datos espaciales se basa en la naturaleza del dominio espacial en estudio. Dependiendo de esto, podemos tener: datos geostadísticos, datos de patrones de puntos y datos lattice (véase Fig. 4.4, tomada de Hengl (2009)).

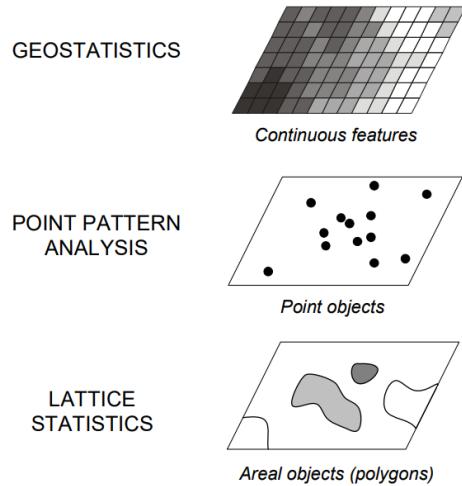


Figura 4.4: Clasificación de datos espaciales propuesta por @cressie1993

GEMA: PUEDES REVISAR LA FORMULA?

Siguiendo a [Cressie \[1993\]](#), sea $s \in \mathbb{R}^d$ una localización en un espacio Euclídeo d -dimensional y $Z(s) : s \in \mathbb{R}^d$ una función aleatoria espacial, donde Z representa el atributo en el cual estamos interesados:

1. **Datos geoestadísticos:** Surgen cuando el dominio de estudio es **continuo y fijo** D . Es decir: (i) $Z(s)$ se puede observar en cualquier punto del dominio (continuo); y (ii) los puntos en D no son estocásticos (son fijos, D es el mismo para todas las realizaciones de la función aleatoria espacial).

Algunos ejemplos de datos geoestadísticos son el nivel de un contaminante en una ciudad, los valores de precipitación o temperatura del aire en un país, las concentraciones de metales pesados en la capa superior del suelo de una región, etc.

Es obvio que, al menos en teoría, el nivel de un contaminante específico podría medirse en cualquier lugar de la ciudad; Lo mismo puede decirse de las mediciones de precipitaciones o temperaturas del aire en un país o concentraciones de un metal pesado en una región. Sin embargo, en la práctica, no es posible una observación exhaustiva del proceso espacial. Por lo general, el proceso espacial se observa en un conjunto de ubicaciones (por ejemplo, el nivel de un contaminante específico en una ciudad se observa en los puntos donde están ubicadas las estaciones de monitoreo) y, basado en tales valores observados, el análisis geoestadístico reproduce el comportamiento de el proceso espacial en todo el dominio de interés.

En el análisis geoestadístico lo más importante es cuantificar la correlación espacial entre observaciones (a través de la herramienta básica en geoestadística, el semivariograma) y utilizar esta información para lograr los objetivos anteriores.

```
# ejemplo cual??? tmin interpolado
```

2. **Datos reticulares:** Surgen cuando: (i) el dominio bajo estudio D es **discreto**, es decir, $Z(s)$ puede observarse en una serie de ubicaciones fijas que pueden enumerarse. Estas ubicaciones pueden ser puntos o regiones, pero generalmente son códigos postales, pistas censales, vecindarios, provincias, países, etc., y los datos en la mayoría de los casos son datos agregados espacialmente sobre estas áreas. Aunque estas regiones pueden tener una forma regular, normalmente la forma que tienen es irregular, y esto, junto con el carácter espacialmente agregado de la datos, es por lo que los datos lattice tambien se denominan datos regionales. Y (ii) las ubicaciones en D no son estocásticas. Por supuesto, un concepto clave en el análisis de los datos lattice es el **vecindario** y la matriz **W**.

Algunos ejemplos de reticulares incluyen la tasa de desempleo por estados, los datos de delincuencia por comarcas, rendimientos agrícolas en parcelas, precios medios de la vivienda por provincias, etc.

ejemplo ¿es el de la renta??

3. **Procesos de puntos:** Mientras que en los datos geoestadísticos y reticulares el dominio D es fijo, en los datos de patrones puntuales el dominio es discreto o continuo, pero **aleatorio**. Los patrones de puntos surgen cuando el atributo bajo estudio es la ubicación de los eventos (observaciones). Es decir, el interés radica en dónde ocurren eventos de interés.

Algunos ejemplos de patrones de puntos son la ubicación de incendios en una región española, la ubicación de los árboles en un bosque o la ubicación de nidos en una colonia de aves reproductoras, la localización de los delitos en una ciudad, entre muchas otras.

En estos casos, es obvio que D es aleatorio y los puntos de observación no dependen del investigador. El principal objetivo del análisis de patrones de puntos es determinar si la ubicación de los eventos tiende a exhibir un patrón sistemático sobre el área en estudio o, por el contrario, son aleatoriamente repartido.

Más concretamente, nos interesa analizar si la ubicación de los eventos es completamente aleatorio espacialmente (la ubicación donde ocurren los eventos no se ve afectada por la ubicación de otros eventos), uniforme o regular (cada punto está tan lejos de todos sus vecinos como sea posible) o agrupados o agregados (la ubicación de los eventos se concentra en grupos).

ejemplo, el de los accidentes??

Apéndice A

Tipos de CRS proyectados

Existen varias familias de proyecciones, que se pueden clasificar de diversas maneras: por tipo de superficie de proyección y por métrica a preservar.

A.1. Por tipo de superficie de proyección

El proceso de trasladar puntos de una esfera a un plano puede plantearse de manera práctica como el ejercicio de envolver una esfera con una superficie plana (como una hoja de papel) y trasladar los puntos de la esfera de manera lineal al punto de la superficie plana más cercano a ella. La Fig. A.1 muestra estos tres tipos de proyección.

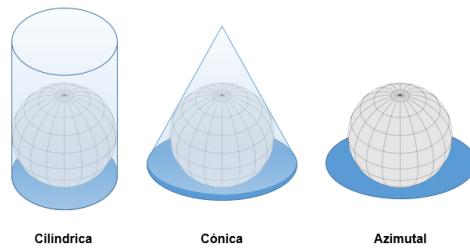


Figura A.1: Tipos de proyección por superficie de proyección

A partir de este ejercicio, se plantean tres posibles soluciones (cilíndrica, cónica y acimutal o planar), dependiendo del tipo de superficie que se use para proyectar.

- **Proyecciones cilíndricas:** Son aquellas proyecciones donde la superficie de proyección conforma un cilindro alrededor de la Tierra. Una de las

proyecciones cilíndricas más conocidas es la **proyección de Mercator** (Ver Fig. A.2).

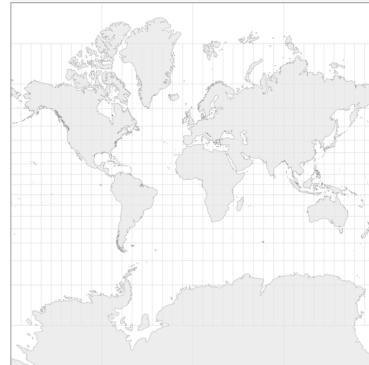


Figura A.2: Proyección Mercator

- **Proyecciones cónicas:** En este tipo de proyecciones, se plantea la superficie de proyección como una forma cónica. Como ejemplo, la **proyección cónica equiáreas de Albers** es una de las proyecciones que más suele usarse en la representación de mapas de América del Norte (Ver Fig. A.3).

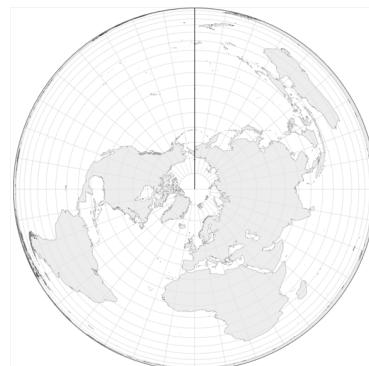


Figura A.3: Proyección cónica equiáreas de Albers

- **Proyecciones acimutales o planares:** En este tipo de proyección se proyecta una porción de la Tierra sobre un plano que es tangente a la misma en el punto de referencia. Como ejemplos de proyecciones acimutales podemos destacar la **proyección ortográfica** (Ver Fig. A.4).



Figura A.4: Proyección ortogonal

A.2. Por métrica a preservar

Es importante tener en cuenta que cualquier proyección de la superficie de la Tierra produce distorsiones en una o varias características geográficas. Como ejemplo clásico, la proyección de Mercator produce distorsiones del **tamaño** especialmente en aquellas regiones más cercanas a los polos (Groenlandia, que la proyección de Mercator presenta una área similar a la de África, tiene menor superficie real que Argelia). Otras de las métricas que suele verse distorsionada son la **distancia** entre dos puntos geográficos, la **dirección** o la **forma** de regiones de la Tierra.

A lo largo de la Historia se han desarrollado diversas proyecciones cuyo objetivo es preservar alguna o varias de las propiedades mencionadas anteriormente, sin embargo es importante destacar que **no existe una proyección que sea capaz de preservar todas las métricas a la vez**.

Según la metrica a presevar, las proyecciones se pueden clasificar en:

- **Proyecciones conformales:** Intentan preservar los **ángulos** que se forman en la superficie terrestre. Por ejemplo, la proyección de Mercator representa ángulos rectos en las intersecciones de los paralelos y los meridianos (Ver Fig. A.5).
- **Proyecciones equivalentes:** Preservan las proporciones de las **áreas**, provocando a su vez deformaciones en el resto de características, como la forma o los ángulos. La proyección acimutal equivalente de Lambers es un tipo de proyección equivalente (Ver Fig. A.6).
- **Proyecciones equidistantes:** Preservan la **distancia** entre dos puntos geográficos específicos. Por ejemplo, la proyección Plate carré preserva la distancia entre el Polo Norte y el Polo Sur (Ver Fig. A.7).

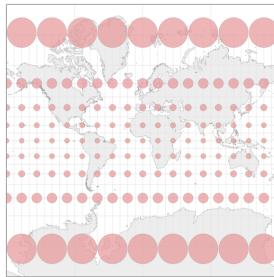


Figura A.5: Ejemplo de proyección conformal: Mercator

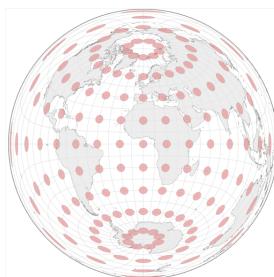


Figura A.6: Ejemplo de proyección equivalente: Proyección acimutal equivalente de Lambers

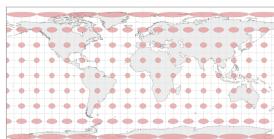


Figura A.7: Ejemplo de proyección equidistante: Platé carre

- **Proyecciones de compromiso:** No intentan preservar ninguna métrica en concreto. En su lugar, se centran en intentar encontrar un **equilibrio** entre las diversas distorsiones que provocan para intentar dar una representación más o menos representativa de la superficie terrestre. La proyección de Winkel Tripel, usada en los mapas de National Geographic, es un ejemplo de proyección de compromiso (Ver Fig. A.8).

En los anteriores ejemplos se ha añadido a cada proyección la **indicatriz de Tissot**. Ésta consiste en una serie de círculos imaginarios de igual área distribuidos sobre la superficie esférica de la Tierra en determinados puntos. De este manera, al presentar la indicatriz de Tissot en una proyección específica, se puede entender de una manera intuitiva la distorsión provocada por dicha proyección, ya que los círculos se ven distorsionados o preservados según los parámetros y la naturaleza de la proyección en cuestión.

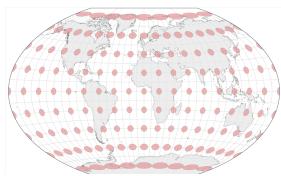


Figura A.8: Ejemplo de proyección de compromiso: Winkel Tripel

Bibliografía

- Robert J. Abrahart, Stan Openshaw, Robert J. Abrahart, and Linda M. See, editors. *Geocomputation*. CRC Press, 05 2000. doi: 10.4324/9780203305805. URL <http://dx.doi.org/10.4324/9780203305805>.
- Noel A. C. Cressie. *Statistics for Spatial Data*. John Wiley and Sons, Inc., 09 1993. doi: 10.1002/9781119115151. URL <http://dx.doi.org/10.1002/9781119115151>.
- Diego Hernangómez. *mapSpain: Administrative Boundaries of Spain*, 2022. URL <https://ropenspain.github.io/mapSpain/>.
- Robin Lovelace, Jakub Nowosad, and Jannes Muenchow. *Geocomputation with R*. R Series. CRC Press, 2019. ISBN 1-138-30451-4.
- J.M. Montero, G. Fernández-Avilés, J. Mateu, and E Porcu. *Geoestadística espacial y espacio-temporal: vino nuevo, cepas viejas*. Riobóo and Riobóo (Eds.) Historia de la Probabilidad y la Estadística, 2011. ISBN 978-84-938774-9-1.
- I Rees, P y Turton. Guest editorial. *Environment and Planning A: Economy and Space*, 30(10):1835–1838, 10 1998. doi: 10.1068/a301835. URL <http://dx.doi.org/10.1068/a301835>.
- Kyle Walker. *crsuggest: Obtain Suggested Coordinate Reference System Information for Spatial Data*, 2021. R package version 0.3.1.