Estadística Espacial

Análisis de datos espaciales: datos que tienen asociados una localización, una posición en el espacio. Los **datos tienen un índice de posición**. Cuando la referencia de posición es geográfica serán datos geoespaciales (datos georeferenciados)

Objetivos del análisis espacial:

- Posicionar y representar en el espacio nuestras variables y entender como se distribuyen en el espacio
  - Distribución de precios en la Ciudad de Madrid
  - Distribución de la población o de la densidad de población
  - Distribución de la renta ¿hay concentración de la renta en determinadas zonas o territorios?.
- Medir áreas y distancias (por ejemplo medir el área afectada por un incendio o un plaga, y analizar cuantas fincas y tipos de cultivo se ha visto afectada)
- Calcular ubicaciones y rutas óptimas (por ejemplo para empresas de reparto)

Estadística Espacial

Análisis de datos espaciales: datos que tienen asociados una localización, una posición en el espacio. Cuando la referencia de posición es geográfica serán datos geoespaciales (datos georeferenciados)

… más objetivos del análisis espacial:

- Buscar patrones: ¿la distribución de la localización es puramente aleatoria o sigue algún patrón?
- Realizar predicciones o interpolaciones (**Geoestadística**: Krigeado y semivariogramas)

  Por ejemplo, estimación de curvas de Nivel: Presión atmosférica o también mapas de precio de la vivienda
- Modelos de regresión espacial: **Correlación Espacial (Econometría Espacial)**

Software para representación y análisis GIS - Sistema Información Geográfica

- QGIS https://qgis.org/es/site/ (código abierto-software libre)
- ARGIS (se puede conectar con R)
- Carto (https://carto.com/)

Software para la estimación de modelos de econometría espacial

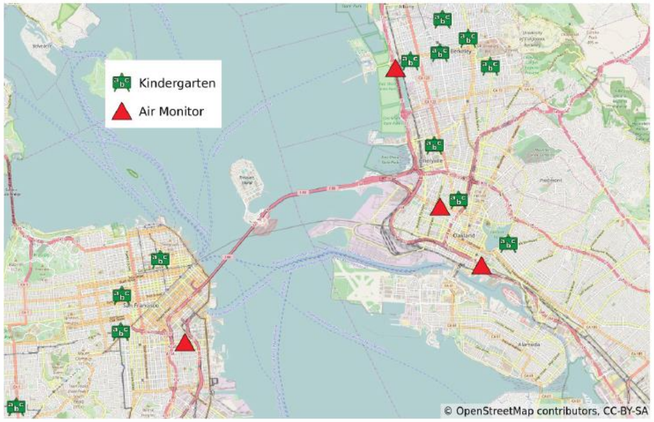
- Stata, SAS, SPSS, python, y … R
- GEODA y Geoda Space

Bibliografía:

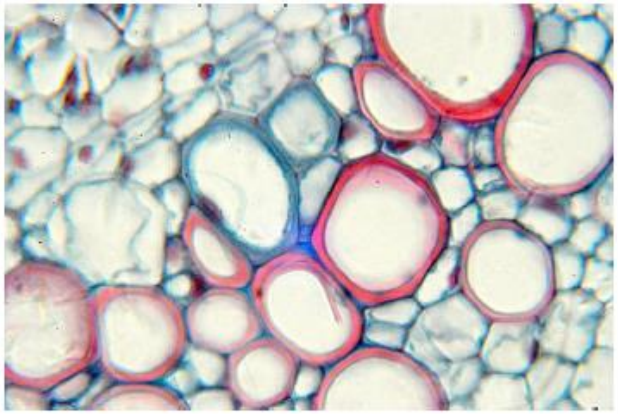
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models. Kluwer Academic Press.*
- Anselin y Rey (2014): *Modern Spatial Econometrics in Practice,* GEODA Press
- Bivand, R.S., E. Pebesma & V. Gómez-Rubio (2013). *Applied Spatial Data Analysis with R, 2nd Edition. Springer.*
- Dubé, J. y Legros, D. (2014): *Spatial Econometrics Using Microdata*. Editorial Iste Ltd and John Wiley & Sons.
- *Elhorst, J.P. (2010). Applied Spatial Econometrics: Raising the Bar, Spatial Economic Analysis, 5(1), 9-28.*
- *Elhorst, J.P. (2014). Spatial Econometrics: From Cross-Sectional Data to spatial Panels. Springer*
- LeSage, J., & Pace, R. K. (2010). *Introduction to Spatial Econometrics*. CRC Press.
- Mas, Jean-François (2013). *Análisis Espacial con R. Usa R como un sistema de Información Geográfica.* European Scientific Institute

Datos Espaciales

Los datos espaciales son datos que tienen asociados una localización, una posición en el espacio. Cuando la referencia de posición es geográfica serán datos geoespaciales (datos georeferenciados)



Datos geostadísticos



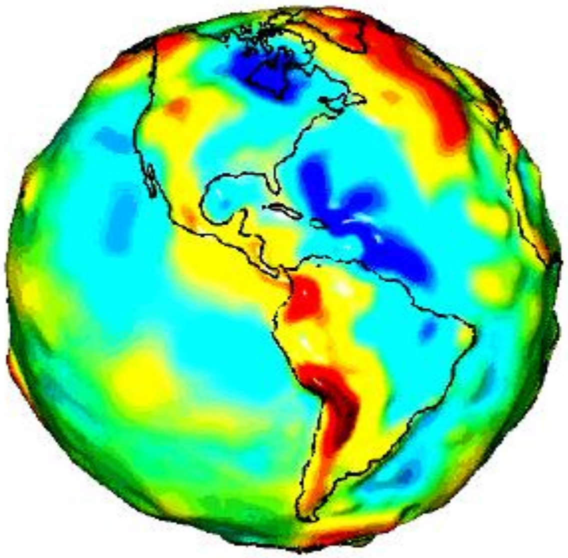
Datos de tejidos celulares

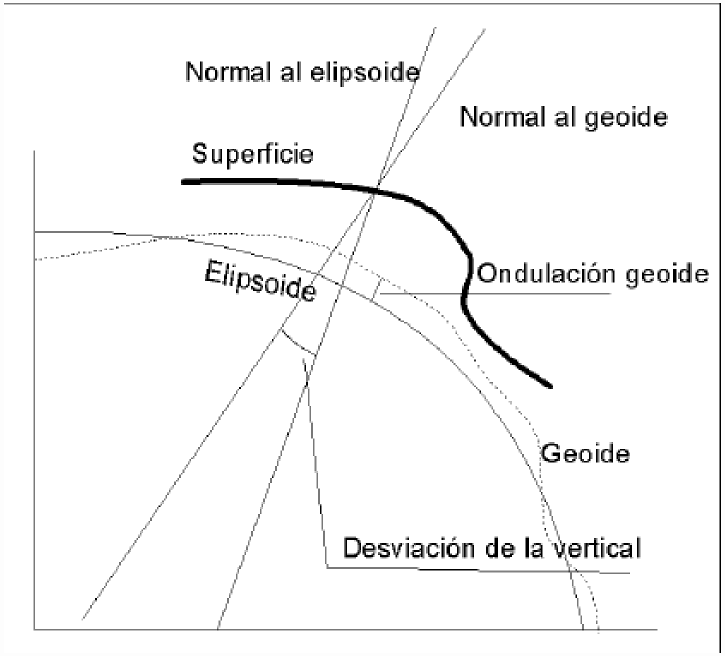
Datos en astronomía

Datos geoespaciales (datos georeferenciados) ¿cómo se asignan coordenadas?

Sistemas de Información Geográfica (GIS)
Proyecciones, Coordenadas, Sistemas de Referencia Espacial, Cartografías …

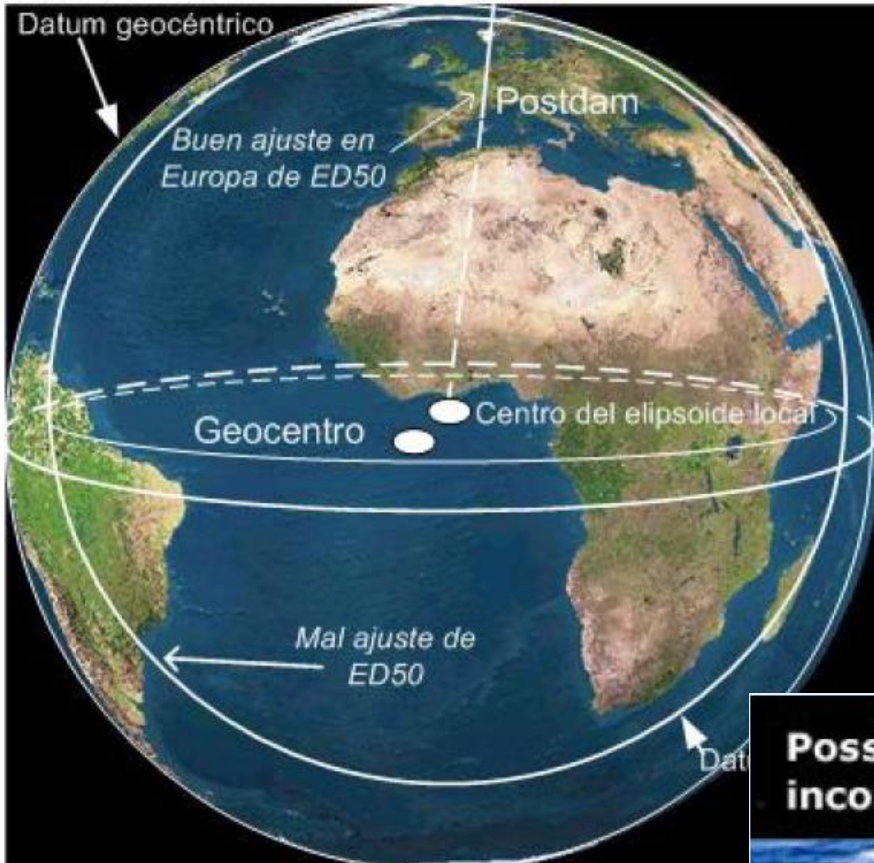
- https://spatialreference.org
- https://www.ign.es/web/ign/portal/gds-area-geodesia
- Ignacio Alonso Fernández- Coppel Localizaciones geográficas.



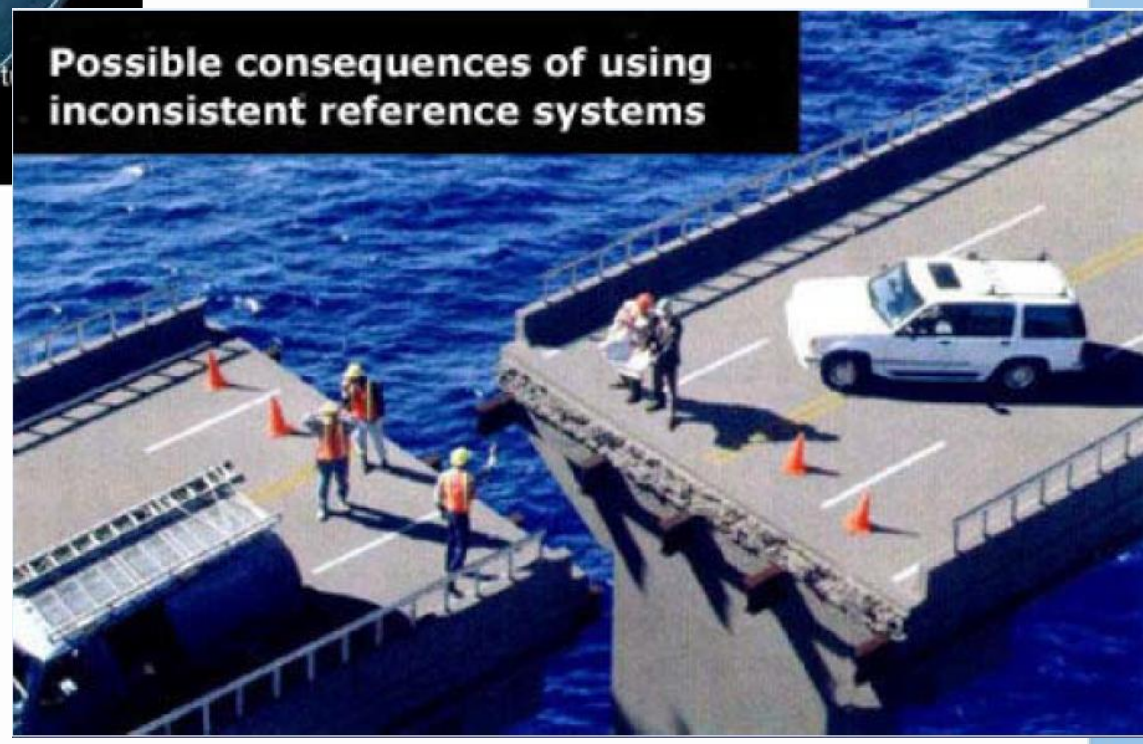


Geoide: superficie (de nivel) equipotencial del campo gravitatorio de la Tierra. Coincide "idealmente" con el nivel medio del mar

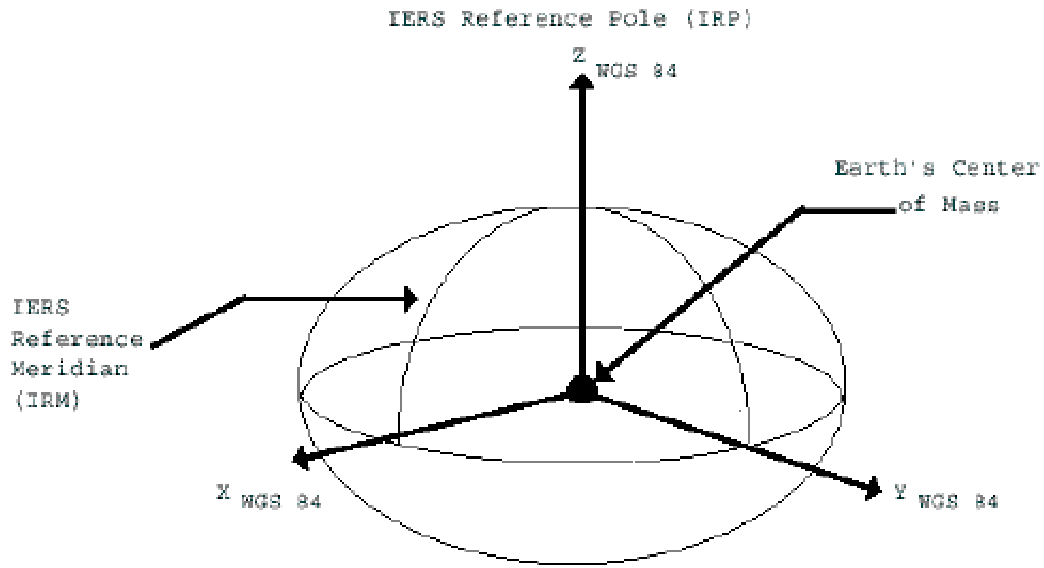
Representación de la superficie de la tierra por la revolución de un elipsoide **que mejor se adapte a la zona concreta** que se quiere representar (**DATUM**: origen y situación de un sistema de coordenadas para una zona de la tierra)



Consecuencia de trabajar con diferentes sistemas geodésicos de referencia



Sistema geodésico de referencia en España ETRS89 (EPSG 25830) (Real Decreto 28/07/07)

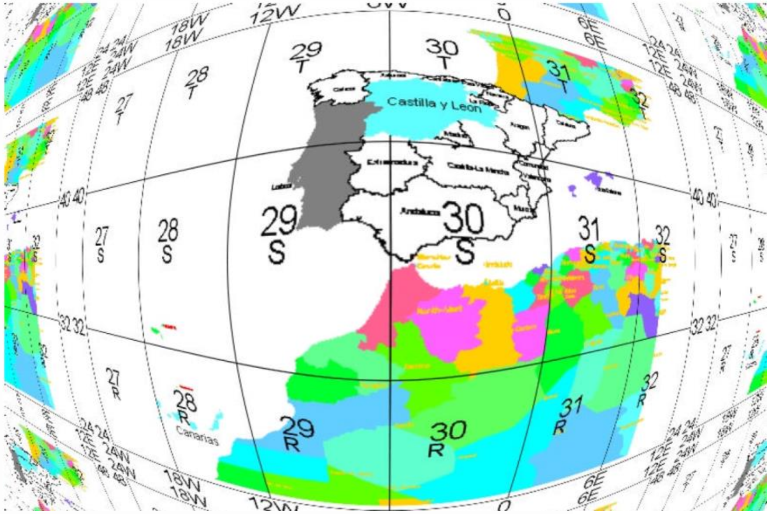


EPSG:25830

ETRS89 / UTM zone 30N (Google it)

- **WGS84 Bounds**: -6.0000, 34.7500, 0.0000, 62.3300
- **Projected Bounds**: 225370.7346, 3849419.9580, 774629.265
- **Scope**: Large and medium scale topographic mapping and engin
- **Last Revised**: Oct. 19, 2000
- **Area**: Europe - 6°W to 0°W and ETRS89 by country

```
PROJCS["ETRS89 / UTM zone 30N",
    GEOGCS["ETRS89",
        DATUM["European_Terrestrial_Reference_System_1989",
            SPHEROID["GRS 1980",6378137,298.257222101,
                AUTHORITY["EPSG","7019"]],
            AUTHORITY["EPSG","6258"]],
        PRIMEM["Greenwich",0,
            AUTHORITY["EPSG","8901"]],
        UNIT["degree",0.01745329251994328,
            AUTHORITY["EPSG","9122"]],
        AUTHORITY["EPSG","4258"]],
    UNIT["metre",1,
        AUTHORITY["EPSG","9001"]],
    PROJECTION["Transverse_Mercator"],
    PARAMETER["latitude_of_origin",0],
    PARAMETER["central_meridian",-3],
    PARAMETER["scale_factor",0.9996],
    PARAMETER["false_easting",500000],
    PARAMETER["false_northing",0],
    AUTHORITY["EPSG","25830"],
    AXIS["Easting",EAST],
    AXIS["Northing",NORTH]]
```



Proyección del elipsoide en un plano

Cartografías para la representación de mapas y datos espaciales
Ejemplo: **archivos.shape**

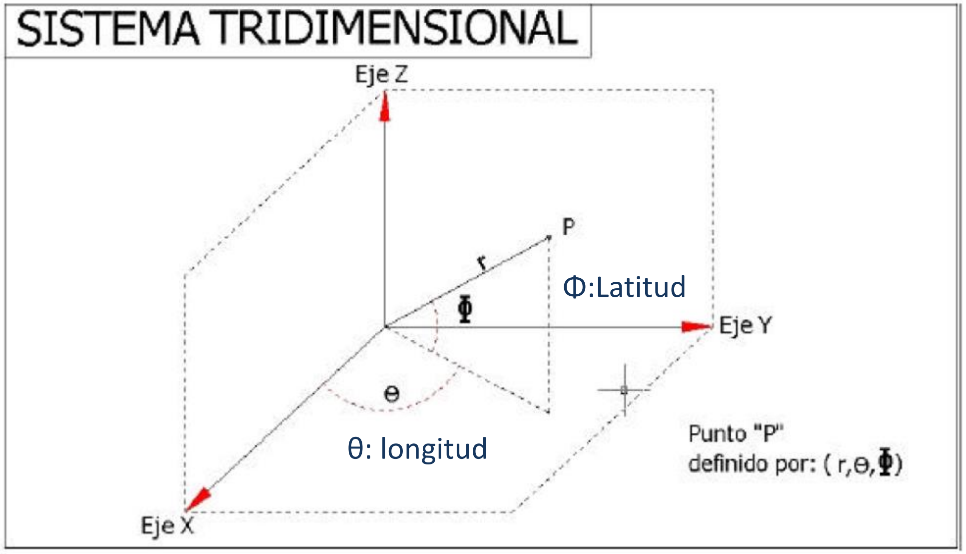
Localización geográfica (coordenadas)

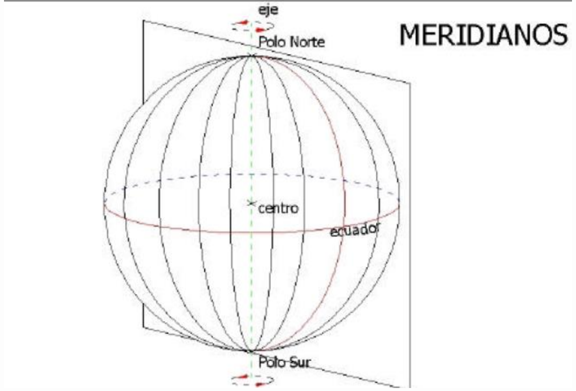
- Coordenadas geográficas en formato Longitud-Latitud
- Coordenadas (x,y) UTM: Universal Transversa Mercator

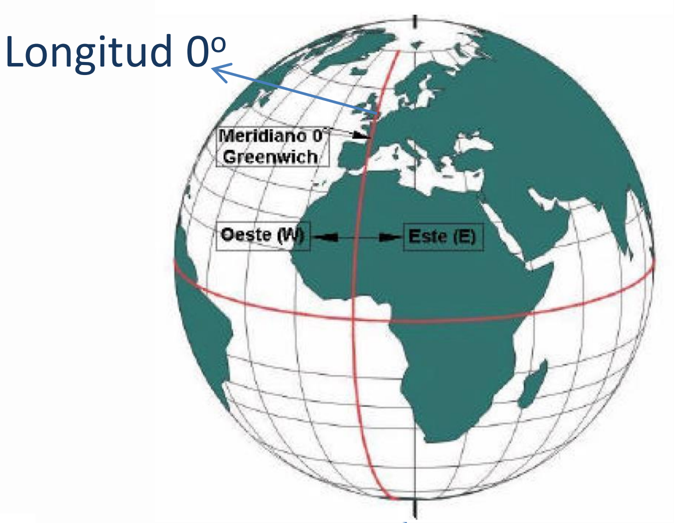
		Coordenada UTM		
Origen	Geográfica	X	y	Origen
Huso 29	42º0'0.0''N4º0'0.0''W	914143.57	4661883.98	Huso 29
Huso 30	42º0'0.0''N4º0'0.0''W	417181.93	4650259.84	Huso 30
Huso 31	42º0'0.0''N4º0'0.0''W	-79874.09	4673541.14	Huso 31

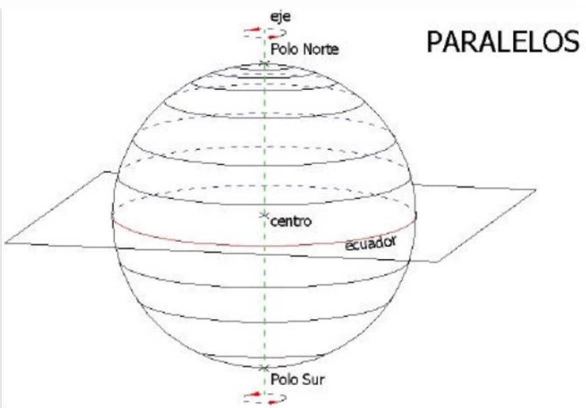
Coordenadas Geográficas

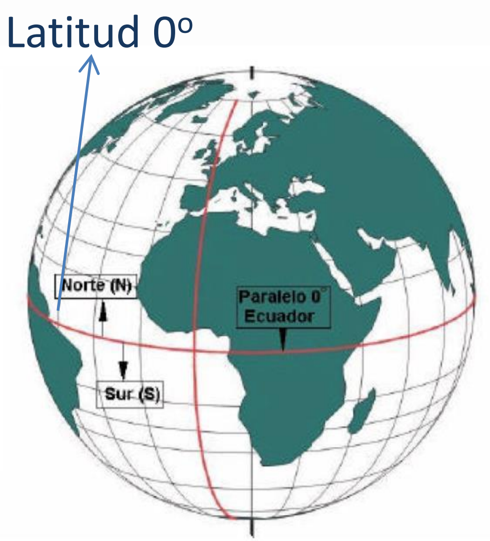
Una forma de designar un punto sobre la superficie terreste:

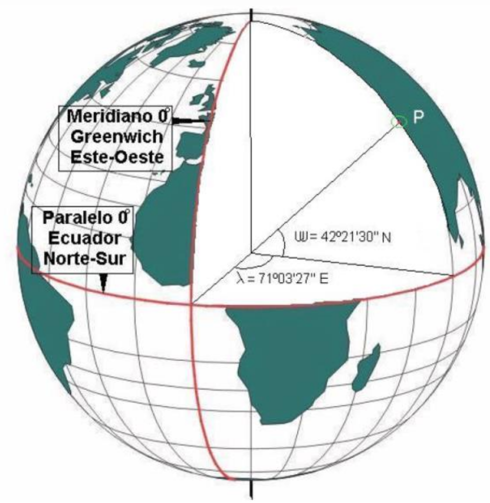






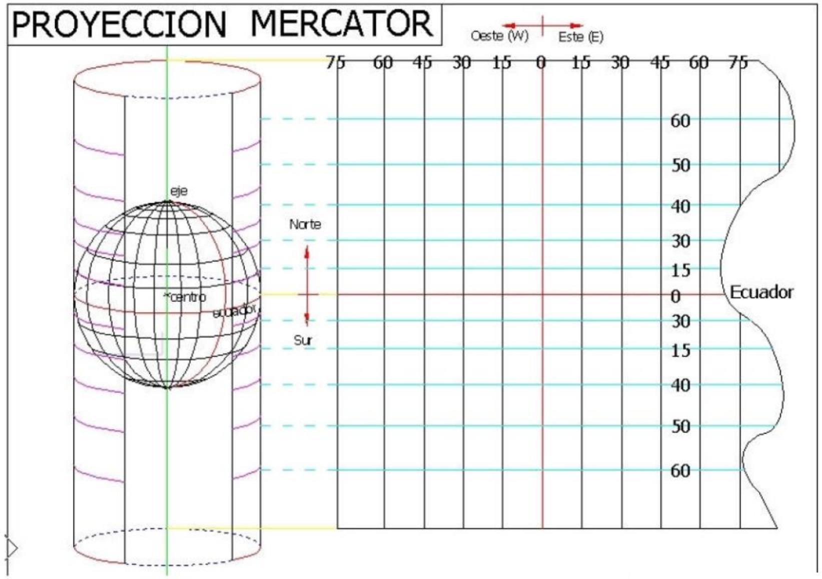


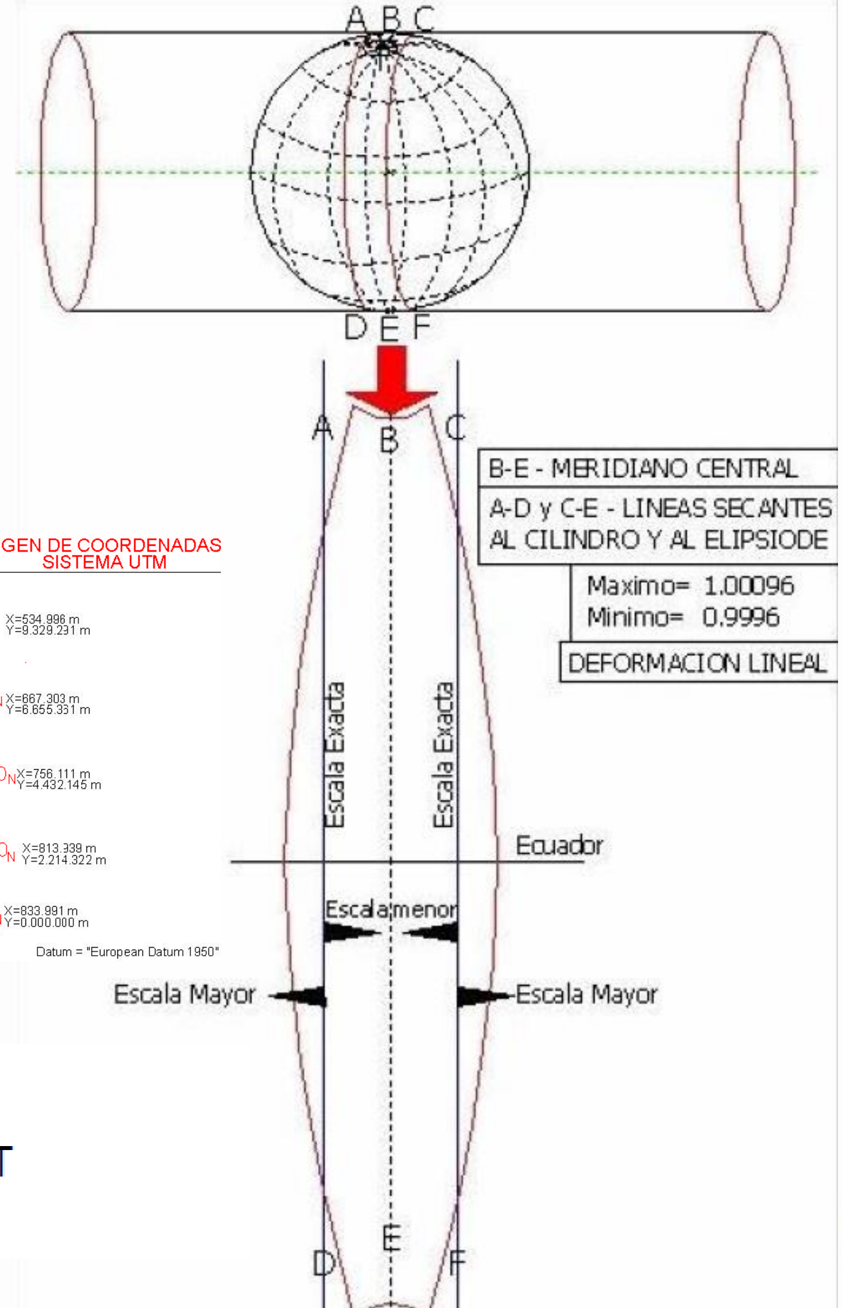


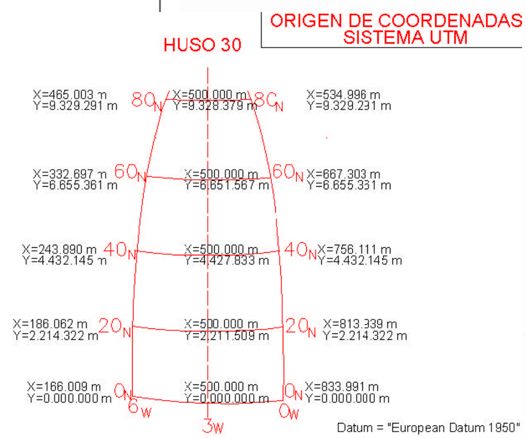


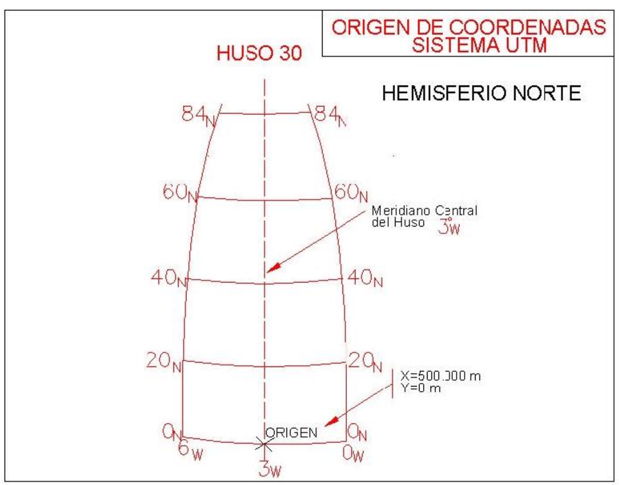
Ignacio Alonso Fernández- Coppel Localizaciones geográficas

UTM: **Universal Transversa Mercator**







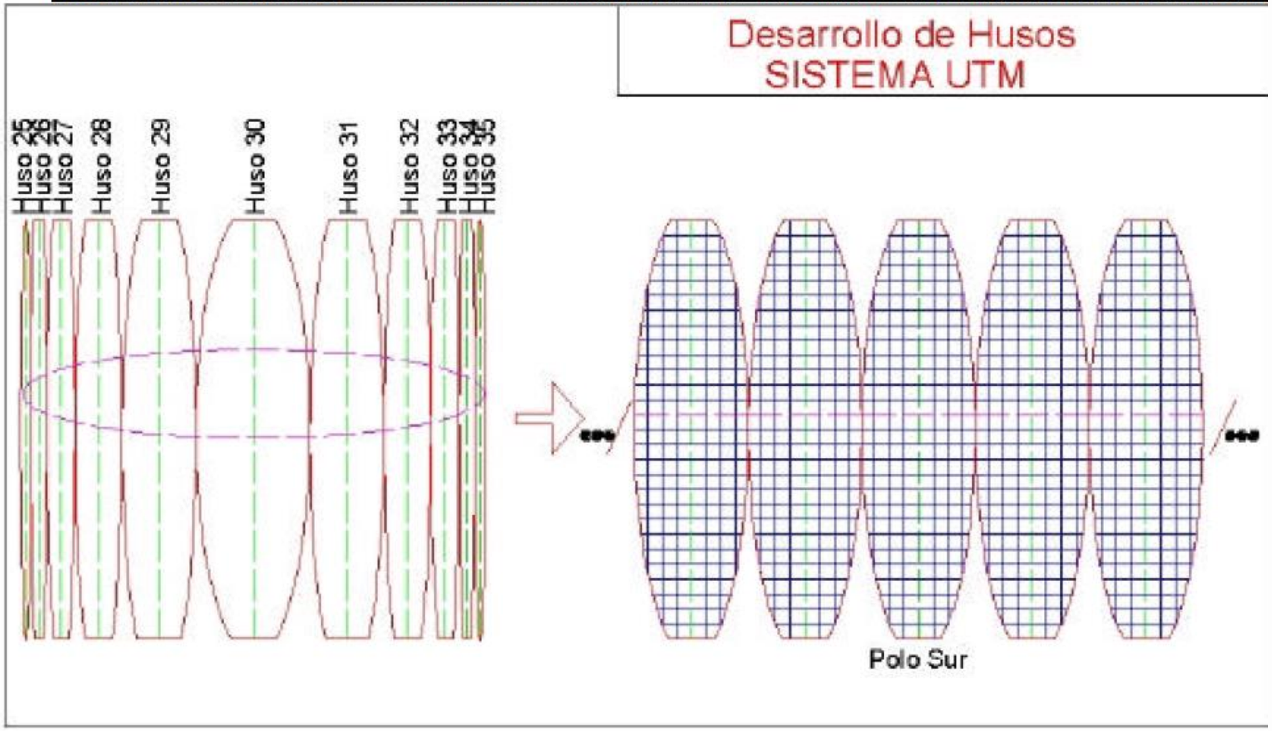


X= 462.130 m

Y= 4.634.140 m

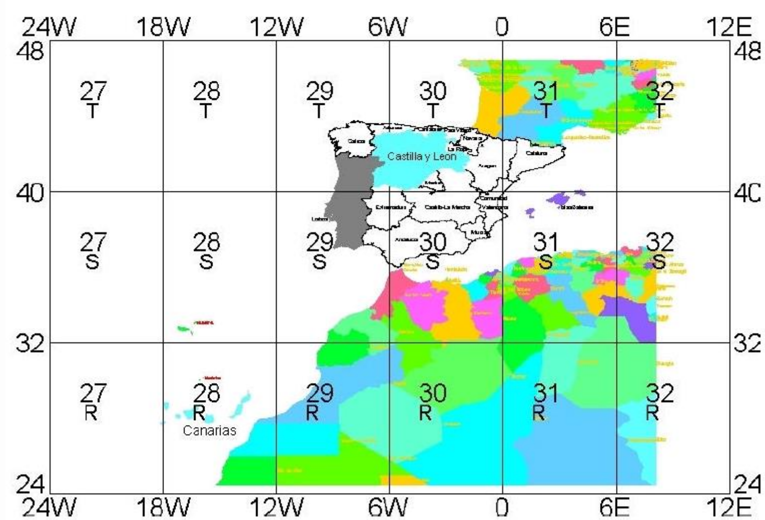
Huso=30 Zona=T

Datum: ETRS89



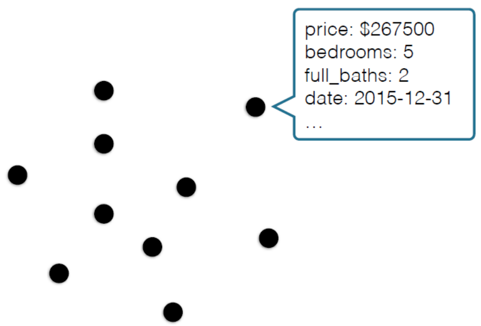
Localizaciones Geográficas. Las Coordenadas Geográficas y la Proyección UTM. (Universal Transversa Mercator)

La distribución de **Husos y zonas** para España es la siguiente:

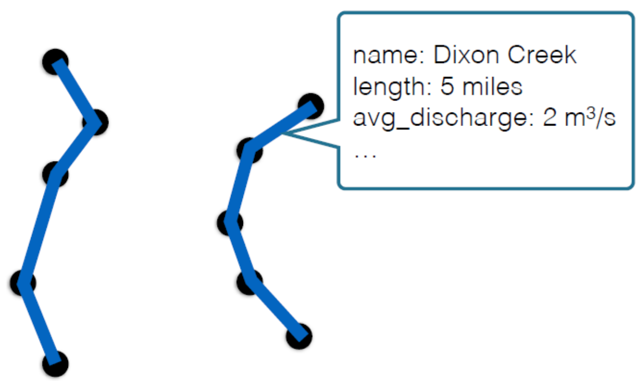


- Tipos de estructura de los datos espaciales

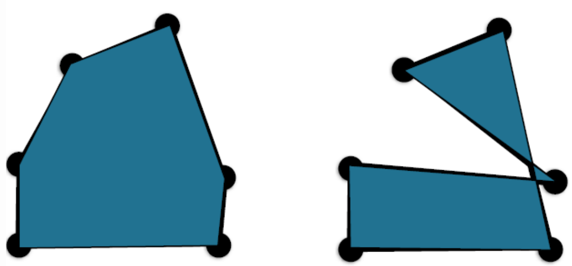
- **Datos vectoriales**: registran la información en forma de coordenadas geográficas (x,y) en un sistema plano de coordenadas. Cada **punto** se registra como una localización simple del tipo (x,y), mientras que las **formas lineales** y los contornos de los **polígonos** se recogen como series ordenadas de coordenadas (x,y).
- **Datos *raster***: registran la información espacial en una cuadrícula regular o matriz organizada en filas y columnas, de forma que cada celda contiene un número que representa una forma geográfica determinada, como un tipo de suelo, la elevación, el uso de la tierra, una pendiente, etc. Ejemplos: píxeles de fotos, mapas marinos de profundidades, etc.



Puntos en el espacio



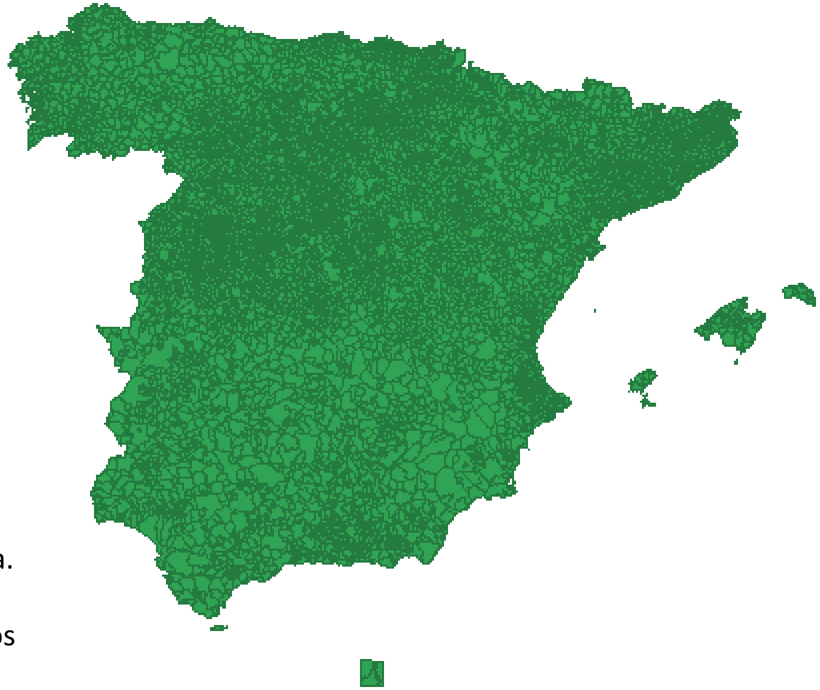
Líneas (carreteras, fronteras)



Polígonos espaciales (municipios, códigos postales, países…)

Cada dato puede llevar asociado valores de las variables correspondientes a cada punto o cada región (se suelen almacenar en una **base de datos .dbf) CLAVES O IDENTIFICADORES**

Para representar los datos espaciales (georoferenciados) en mapas geográficos se necesita disponer de las cartografías



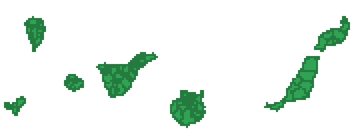
Fuente para obtener cartografías

https://www.naturalearthdata.com/

Existen diferentes formatos de archivos de información geográfica.
El formato mas habitual en econometría espacial es el SHP.
Este tipo de formato es multiarchivo, es decir consiste en al menos tres archivos relacionados:

- .shp: archivo de formas geométricas.
- .shx: archivo índice de las formas geométricas.
- .dbf: archivo de atributos (base de datos).

El formato SHP puede ser directamente ledo por cualquier programa GIS.



El detalle y precisión de la cartografía dependerá de su escala. Dependiendo de para qué queramos la cartografía necesitaremos mayor o menor nivel de detalle. Ejemplo [Eurostat]

Marco Geoestadístico Nacional de México

Divisiones Geoestadísticas de México

AGEE
Área Geoestadística Estatal

AGEM
Área Geoestadística Municipal

AGEB RURAL
Área Geoestadística Básica RURAL

Localidad
Localidad Rural

Manzana
Manzana Rural

Localidad URBANA
Localidad Urbana

AGEB URBANA
Area Geoestadística Básica Urbana

MANZANA
Manzana

Área Geoestadística Estatal (AGEE)

Corresponde al área geográfica de cada una de las 32 entidades federativas del país (31 estados y un Distrito Federal, conformando un total de 32 AGEE).

Se codifica o clavifica de acuerdo con el orden alfabético de sus nombres oficiales, con una longitud de dos dígitos

CVE_ENT	NOM_ENT
01	Aguascalientes
02	Baja California
03	Baja California Sur
04	Campeche
05	Coahuila de Zaragoza
06	Colima
07	Chiapas
08	Chihuahua
09	Distrito Federal
10	Durango
11	Guanajuato
12	Guerrero
13	Hidalgo
14	Jalisco
15	México
16	Michoacán de Ocampo
17	Morelos
18	Nayarit
19	Nuevo León
20	Oaxaca
21	Puebla
22	Querétaro
23	Quintana Roo
24	San Luis Potosí
25	Sinaloa
26	Sonora
27	Tabasco
28	Tamaulipas
29	Tlaxcala
30	Veracruz de Ignacio de la Llave
31	Yucatán
32	Zacatecas

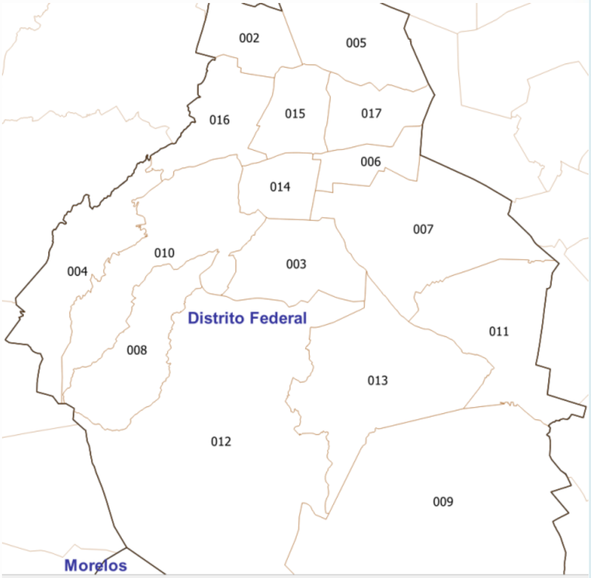
Clave completa o concatenada desde estado hasta manzana	
Para las áreas urbanas: • EE+MMM+LLLL+AAA-A+NNN.	Para las áreas rurales: • EE+MMM+AAA-A+LLLL+NNN.
Donde: • EE = Estado (se representa con dos dígitos, 00). • MMM = Municipio (se representa con tres dígitos, 000). • LLLL = Localidad (se representa con cuatro dígitos, 0000). • AAA-A = Ageb (se representa con tres dígitos, un guión y un dígito verificador, 000-0). • NNN = Manzana (se representa con tres dígitos, 000).	

Fuente: Manual de carografía geostadística. INEGI

Área Geoestadística Municipal (AGEM)

Es la extensión territorial que corresponde al espacio geográfico de cada uno de los municipios que conforman la división política de las entidades federativas de los Estados Unidos Mexicanos; el número total de las AGEM por estado será igual al total de sus municipios; y en el caso del Distrito Federal, son las delegaciones políticas; actualmente existen a nivel nacional 2 456 municipios.

Dentro de estas áreas se encuentran todas las localidades urbanas y rurales que pertenecen a cada uno de los municipios y delegaciones



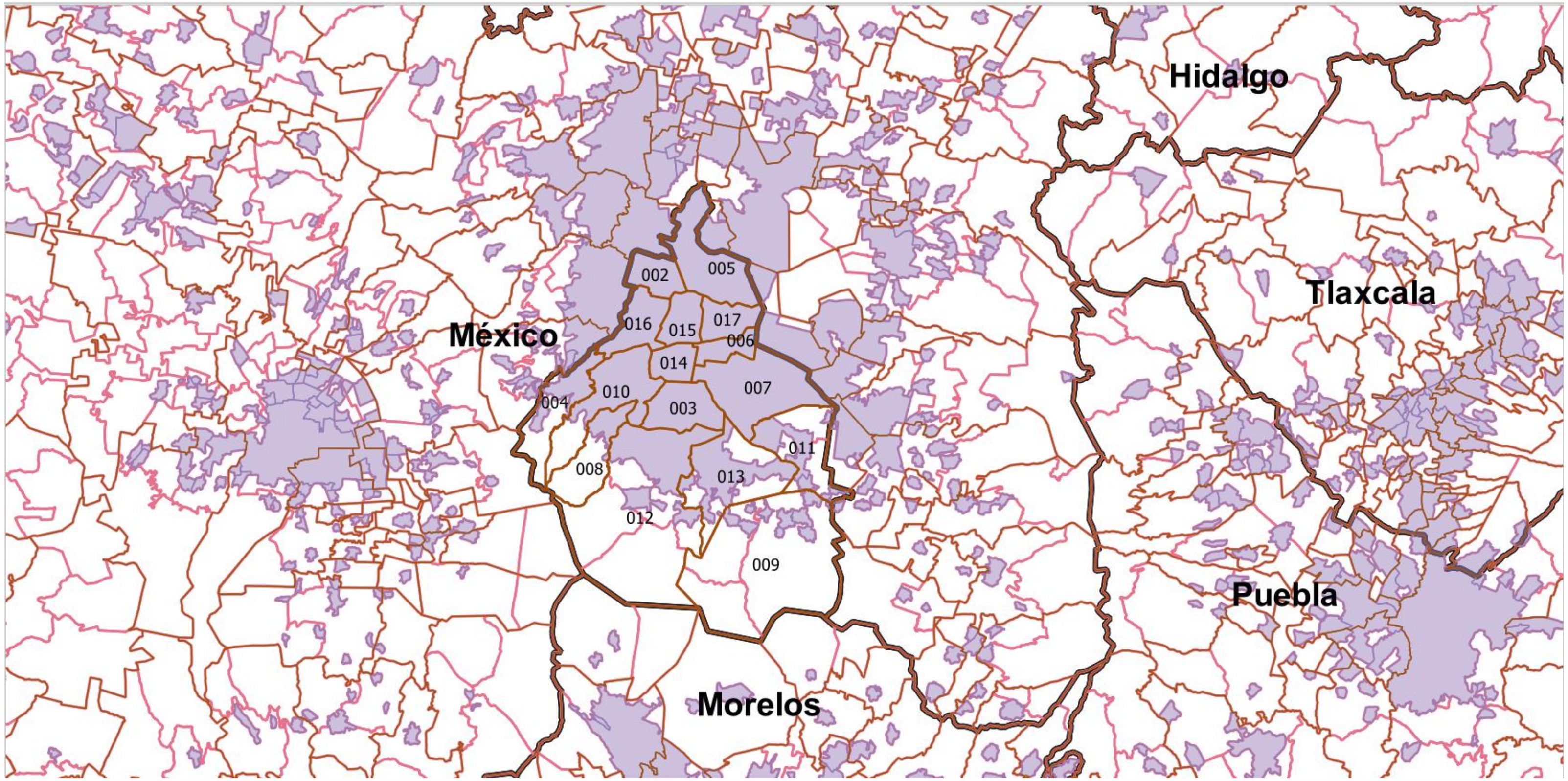
a clave geoestadística de estas áreas está formada por tres números que se asignan de manera ascendente a partir del 001 dentro de cada Entidad

Área Geoestadística Básica (AGEB)

Espacio geográfico menor al municipio, de superficie variable, que constituye la unidad básica del Marco Geoestadístico Nacional. Es la extensión territorial que corresponde a la subdivisión de las áreas geoestadísticas municipales. Constituye la unidad básica del Marco Geoestadístico Nacional y, dependiendo de sus características, se clasifican en dos tipos:

- Área geoestadística básica urbana: Incluye localidades.
- Área geoestadística básica rural.

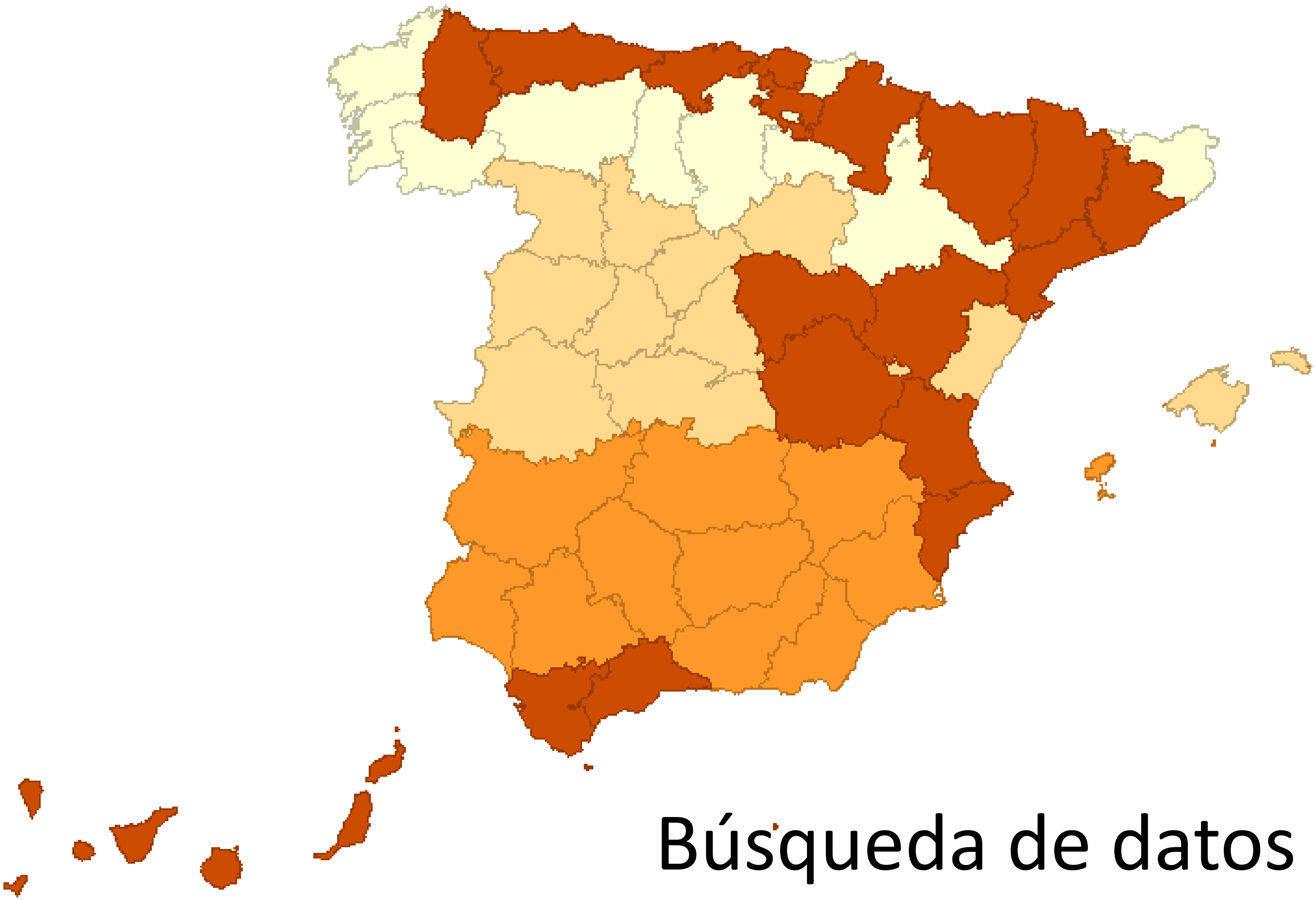
La clave geoestadística de estas AGEB están formadas por tres números un guion y un número que va del 0 al 9 o la letra A dentro de cada Municipio



Selección Area Metrololitana del Valle de Mexico

"CVE_ENT"='09' OR
("CVE_ENT"='13'
AND "CVE_MUN"
='069') OR
("CVE_ENT"='15'
AND ("CVE_MUN" IN
('002','009','010','011'
,'013','015',
'016','017','020','022',
'023','024','025','028',
'029','030','031','033',
'034','035','036','037',
'038','039','044','046',
'050','053','057','058',
'059','060','061','068',
'069','070','075','081',
'083','084','089','091',
'093','094','095','096',
'099','100','103','104',
'108','109','112','120',
'121','122','125',
'065','092')))

Choropleth maps: representación de los valores de una variable estadística en el mapa



Búsqueda de datos Georeferenciados: http://opendata.esri.es/

Econometría Espacial

Ley Waldo de Tobler (1979) o "Primera Ley de la Geografía": "…todo tiene que ver con todo, pero las cosas cercanas están más relacionadas entre sí que las cosas lejanas".

Waldo R. Tobler (1930-2018)

Modelos de Econometría Espacial, son modelos de regresión donde se tiene en cuenta de manera explícita la **dependencia espacial** o dependencia en el espacio (lo que ocurre en un area, una región o un punto determinado del **espacio está relacionado con lo que ocurre a su alrededor)**

Antecedentes:

- Moran (1948) y Geary (1954) para el cálculo de la autocorrelación espacial
- Los primeros trabajos de estadística espacial de los geógrafos **Cliff y Ord (1972, 1973)**
- **Paelinck y Klaasen (1979)** bautizaron estas técnicas con el nombre explícito de "econometría espacial". Después los trabajos de Cliff y Ord (1981), Anselin (1980, 1988A) y Arbia (1989) impulsan y afianzan esta rama de la econometría.

Análisis exploratorio de los datos espaciales

El análisis exploratorio de datos espaciales (AEDE) constituye la etapa inicial de cualquier estudio econométrico que utilice datos espaciales. Está ideado para el estudio univariante.

Se trata de un conjunto de técnicas (gráficas y numéricas) que permiten:

- Describir y visualizar distribuciones espaciales
- Identificar localizaciones espaciales atípicas (*outliers* espaciales)
- Descubrir patrones de asociación espacial, clústeres (concentración o aglomeración) o puntos calientes/fríos
- Sugerir regímenes **espaciales** u otras formas de heterogeneidad espacial
- Búsqueda de dependencias Espaciales ← **Ley de Tobler**

→ **EFECTOS ESPACIALES**

Efectos espaciales

Distinguimos dos tipos de efectos espaciales: Heterogeneidad espacial y autocorrelación (dependencia) espacial

Heterogeneidad espacial (HE): no homogeneidad del mismo fenómeno analizado en distintos puntos del espacio (una relación entre variables dependientes y variables explicativas), de forma que existe HE cuando exista variación en las relaciones en el espacio

Fuentes HE: Aparece cuando en la estimación del modelo de regresión se utilizan datos de unidades espaciales muy diferentes entre sí. Esto implicará que, en los modelos espaciales, las formas funcionales y los parámetros variarán con la localización geográfica no siendo homogéneos para toda la matriz de datos.

- Inestabilidad estructural: ausencia de *estabilidad en el espacio* del comportamiento de la variable estudiada que lleva a que la forma funcional y los parámetros de una regresión varíen según su localización (no son homogéneos para toda la muestra). Las relaciones en sí pueden ser diferentes a través del espacio.

$$y_i = x_i'\beta_i + u_i$$

- Heterocedasticidad: omisión de variables relevantes que llevan a errores de medida en la especificación que se concretan en varianzas diferentes en el término de error.

$$Var(u_i) = \sigma_i^2$$

Puede tratarse con técnicas econométricas estándar.

Efectos espaciales

Autocorrelación (dependencia) espacial (AE): Los datos observados en un punto del espacio no son independientes de los valores observados en puntos de su alrededor.

Esto es, una variable se encontrará ***espacialmente autocorrelacionada*** cuando los valores observados en un lugar determinado dependen, no sólo de ciertos factores externos (otras variables), sino también de los valores observados en regiones/puntos vecinas ("Primera Ley de la Geografía").

Por ejemplo Cox (1969) demostró que existía una *continuidad geográfica* en las intenciones de voto de los norteamericanos, es decir, que si un estado votaba mayoritariamente al partido demócrata, se producía un aumento en la probabilidad de que los estados vecinos también votaran a los demócratas. Otro ejemplo Cliff y Ord (1981): presenta el análisis espacial como forma de conocimiento de las causas y formas de propagación de epidemias y enfermedades.

Fuentes de la Autocorrelación Espacial:

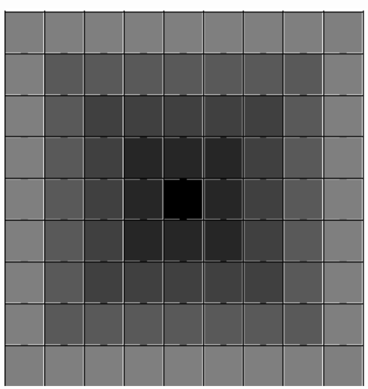
- Errores de medición para las observaciones en unidades espaciales contiguas por delimitación arbitraria de las unidades espaciales (que no recogen bien el proceso generador de datos).
- Agregación espacial de los datos (que puede no corresponder con su proceso generador).
- Interacción espacial de las unidades: externalidades en forma de *spillovers* (efectos desbordamiento) espaciales y jerarquías espaciales.

Autocorrelación espacial

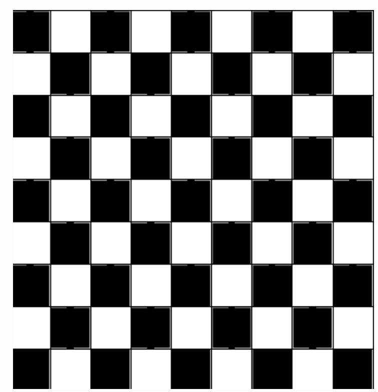
La autocorrelación espacial puede ser positiva o negativa.

- AE **positiva**: la presencia de un fenómeno en una unidad geográfica aumenta significativamente las posibilidades de que ese mismo fenómeno ocurra en la unidades colindantes. Concentración del fenómeno.
- AE **negativa:** la presencia de un fenómeno en una unidad geográfica imposibilita o disminuye significativamente las posibilidades de que ese mismo fenómeno ocurra en las unidades colindantes. Unidades geográficas cercanas son más disímiles entre ellas que entre unidades alejadas en el espacio.
- AE **nula**: la variable se distribuye de forma aleatoria (no existen patrones en la distribución espacial de los valores del fenómeno estudiado).

Podemos interpretar la AE como la **concentración** (positiva) o **dispersión** (negativa) de los valores de una variable en un mapa.



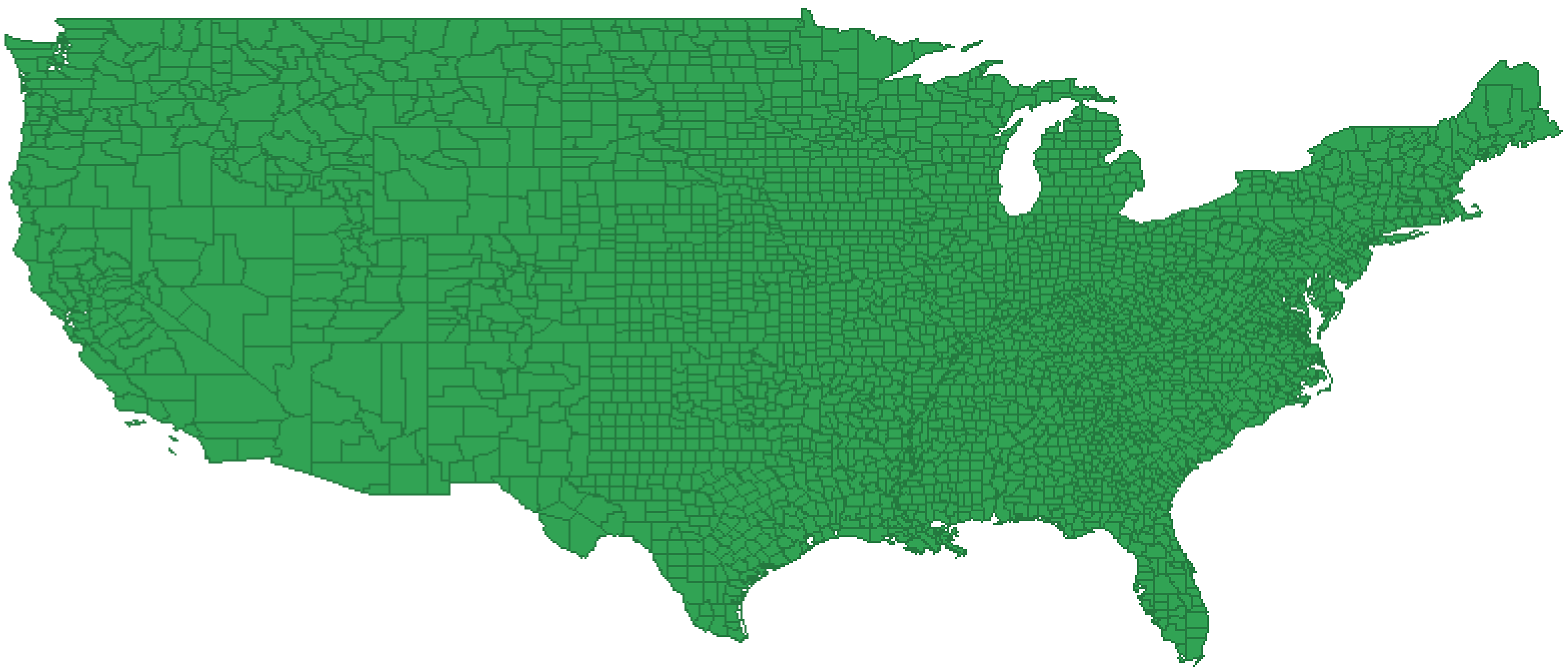
AE positiva



AE negativa

Fuente Gráficos : Baronio, A., Vianco, A. , Rabanal, C. (2012): Una introducción a la econometría espacial

Homicides and selected socio-economic characteristics for continental U.S. counties. Data for four decennial census years: 1960, 1970, 1980, 1990.
Type = polygon shape file, unprojected, lat-lon
Observations = 3085
Variables = 69



Variable: Description

REGIONS: regions (South, West and Midwest/NE):

NOSOUTH: South region dummy

POLYID: unique numeric ID

NAME: county name

STATE_NAME: state name

STATE_FIPS: state fips code (character)

CNTY_FIPS: county fips code (character)

FIPS: combined state and county fips code (character)

STFIPS: state fips code (numeric)

COFIPS: county fips code (numeric)

FIPSNO: fips code as numeric variable

SOUTH: dummy variable for Southern counties (South = 1)

HR**: homicide rate per 100,000 (1960, 1970, 1980, 1990)

HC**: homicide count, three year average centered on 1960, 1970, 1980, 1990

PO**: county population, 1960, 1970, 1980, 1990

RD**: resource deprivation 1960, 1970, 1980, 1990 (principal component, see Codebook for details)

PS**: population structure 1960, 1970, 1980, 1990 (principal component, see Codebook for details)

UE**: unemployment rate 1960, 1970, 1980, 1990

DV**: divorce rate 1960, 1970, 1980, 1990 (% males over 14 divorced)

MA**: median age 1960, 1970, 1980, 1990

POL**: log of population 1960, 1970, 1980, 1990

DNL**: log of population density 1960, 1970, 1980, 1990

MFIL**: log of median family income 1960, 1970, 1980, 1990

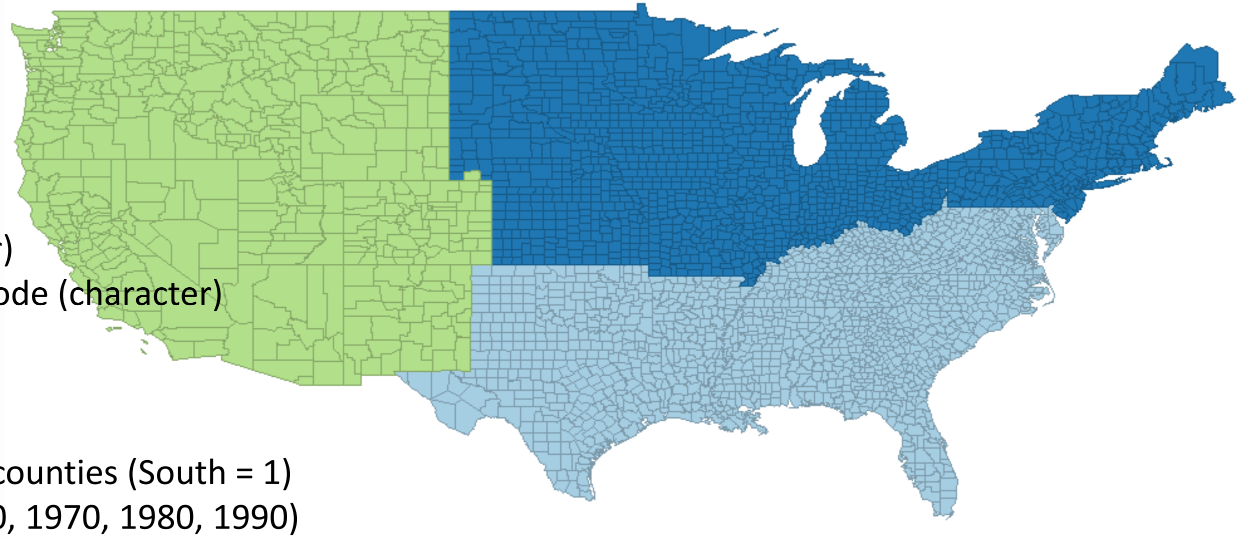
FP**: % families below poverty 1960, 1970, 1980, 1990 (see Codebook for details)

BLK**: % black 1960, 1970, 1980, 1990

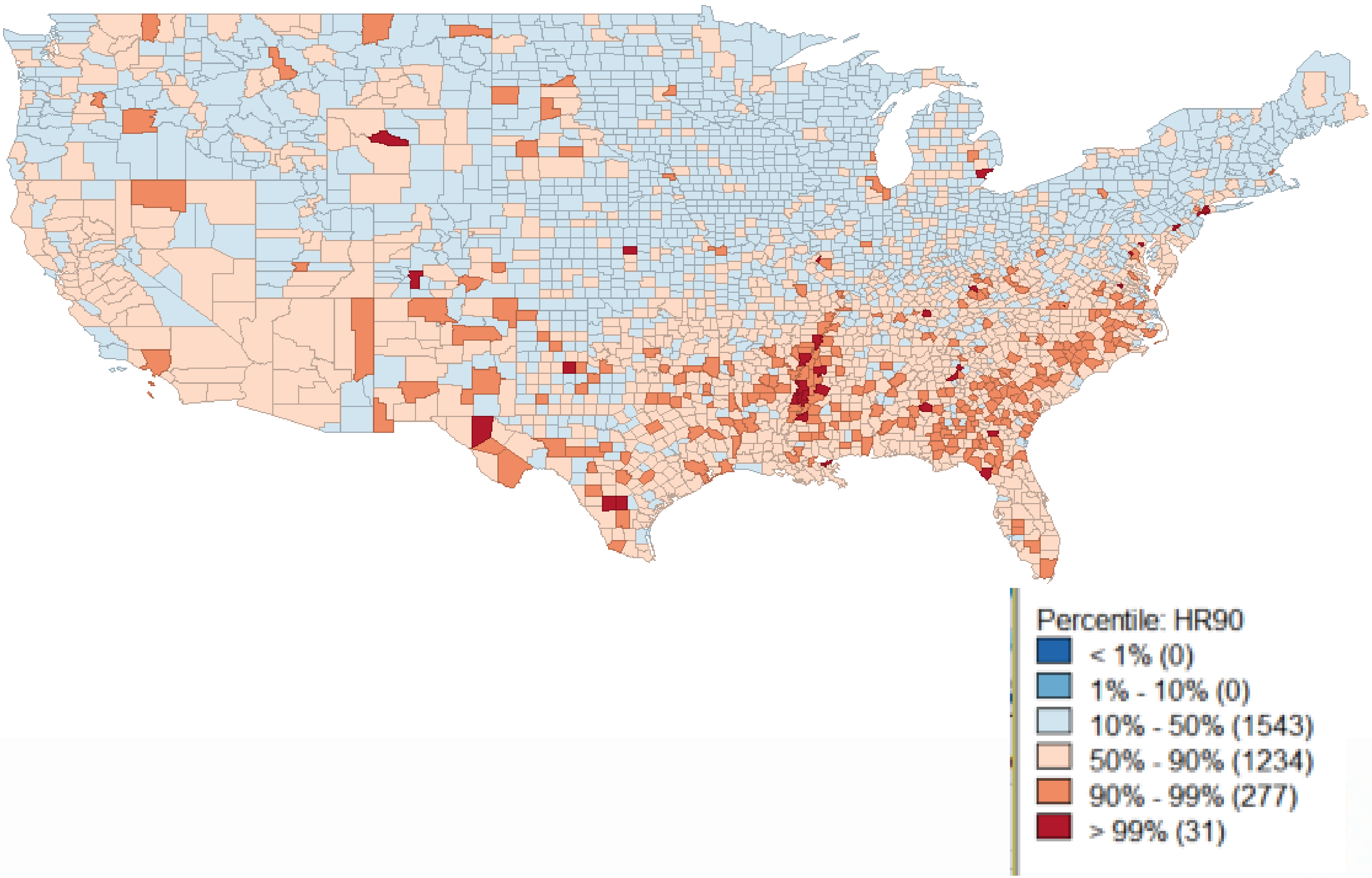
GI**: Gini index of family income inequality 1960, 1970, 1980, 1990

FH**: % female headed households 1960, 1970, 1980, 1990

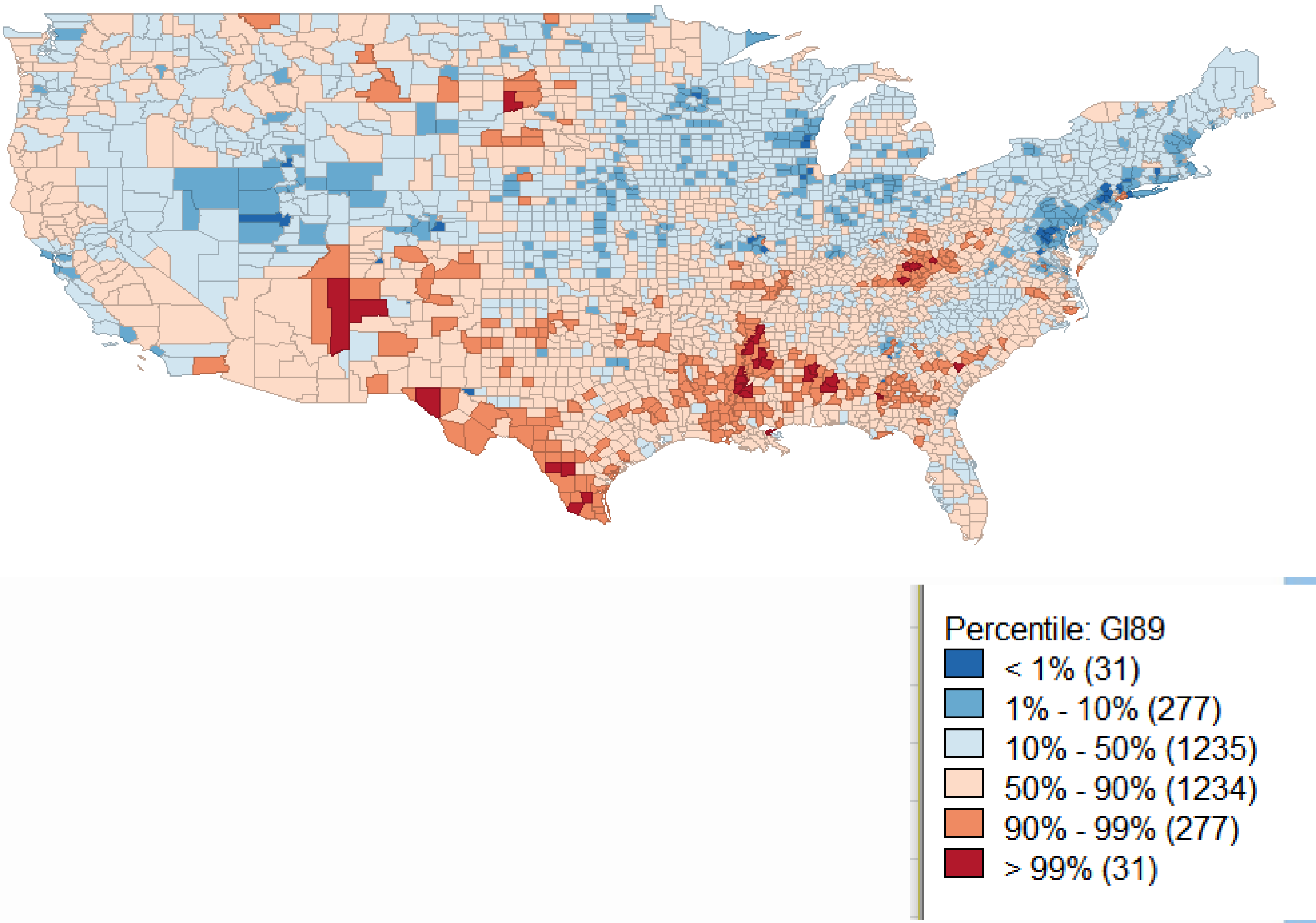
WEST: West region dummy



Homicide rate per 100,000 (1990)



Gini index of family income inequality in 1990



Estimación de la Autocorrelación Espacial:
Matrices de pesos Espaciales por Contigüidad o por distancia

La **dependencia espacial está determinada por una noción de espacio relativo o de localización relativa** que realza el efecto de la **distancia**.

Recordemos que una variable se encontrará *espacialmente autocorrelacionada* cuando los valores observados en un lugar determinado dependen, no sólo de ciertos factores externos (otras variables), sino de los valores observados en regiones/puntos vecinas. Por lo tanto, para estimar la **autocorrelación espacial** es necesario estimar la dependencia entre el valor de una variable en de cada punto o región del espacio con el valor de esa misma variable para sus puntos o **regiones vecinas**

En los **modelos de regresión espacial** se consideran de forma explícita las relaciones de dependencia espacial. El valor de la variable dependiente en cada punto o región dependerá del valor que tomen el resto de variables explicativas en ese punto o región, pero también de lo que esté pasando alrededor, en puntos cercanos o **regiones vecinas**.

¿**Cómo definimos la vecindad**? Existen muchas formas de definir vecindad y distancia.

Nota: Vecindad no implica necesariamente contigüidad física (en el espacio), sino que existen una gran cantidad de criterios para definirla (por ejemplo pueden definirse distancias en términos de volumen de comercio entre países)

Matriz W de pesos espaciales

Para estimar autocorrelación espacial lo primero que hay que determinar son los puntos que están cercanos o las regiones (polígonos) que son colindantes (vecinos). Esto se hace con la ayuda de las **matrices de pesos espaciales** o matriz proximidad espacial, o **matriz de contigüidad o de distancia**, y que se simboliza con **W.**

La matriz W de pesos espaciales es una matriz cuadrada (*nxn*), no estocástica y cuyos pesos w_{ij} **reflejan la intensidad de la interdependencia** en el espacio entre cada par de regiones *i* y *j* *(i, j =1, 2…. n)* .

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix}$$

Dos tipos de matrices de pesos espaciales:

- Para datos de polígonos o regiones: matrices de **contigüidad** o vecindad
- Para puntos (x,y): matrices de **distancia**

La matriz W de pesos espaciales constituye el punto débil de la econometría espacial dado que la robustez de los resultados depende de una correcta elección de la misma.

Matriz W de pesos espaciales: CONTIGÜIDAD

La matriz W de pesos espaciales es una matriz cuadrada, no estocástica y cuyos pesos w_{ij} reflejan la intensidad de la interdependencia en el espacio entre cada par de regiones *i* y *j*.

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix}$$

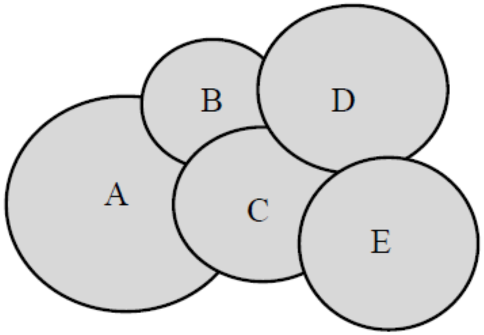
¿Cómo definir los pesos espaciales?

Aunque los pesos han de ser *finitos* y *no negativos*, no existe una definición unánimemente aceptada, se recurre al concepto de *contigüidad física de primer orden*.

Los pesos espaciales:

$$w_{ij} = \begin{cases} 1 & cuando\ (i,j)\ SÍ\ son\ vecinos\ (contigüos) \\ 0 & cuando\ (i,j)\ NO\ son\ vecinos\ (no\ están\ próximos)\ o\ i=j\ (w_{ii}=0) \end{cases}$$

Matriz W de pesos espaciales: CONTIGÜIDAD



	A	B	C	D	E
A	0	1	1	0	0
B	1	0	1	1	0
C	1	1	0	1	1
D	0	1	1	0	
E	0	0	1	1	0

La forma más sencilla de construir una matriz de contigüidad es utilizando la notación binaria:

- 1 representa contigüidad espacial entre dos unidades
- 0 ausencia de contigüidad espacial entre dos unidades
- Los elementos de la diagonal principal son cero, pues ninguna región puede ser vecina de sí misma.

Una matriz construida de esta manera es **simétrica**.

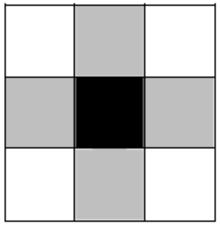
Fuente Gráficos : Baronio, A., Vianco, A. , Rabanal, C. (2012): Una introducción a la econometría espacial

Criterios de contigüidad de primer orden

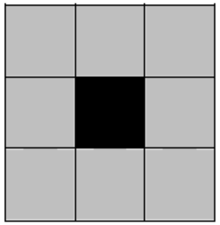
Existen muchas formas de definir la contigüidad (o vecindad) espacial:

- Contigüidad de **torre** (la más usada): w_{ij}=1 para unidades que comparten un lado común con la región de interés y 0 en caso contrario.
- Contigüidad de **alfil**: w_{ij}=1 para unidades que comparten un vértice en común con la región de interés y 0 en caso contrario.
- Contigüidad de **reina**: w_{ij}=1 para unidades que comparten un lado en común o un vértice con la región de interés y 0 en caso contrario.

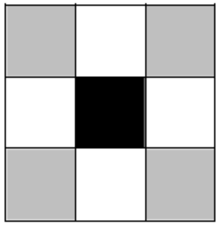
Usando alguno de estos criterios de contigüidad física se construye la matriz de pesos de **primer orden**. Solo es necesario establecer la localización de las unidades en el mapa que tienen bordes comunes con longitudes positivas.



Matriz torre (rook) de primer orden



Matriz reina (queen) de primer orden



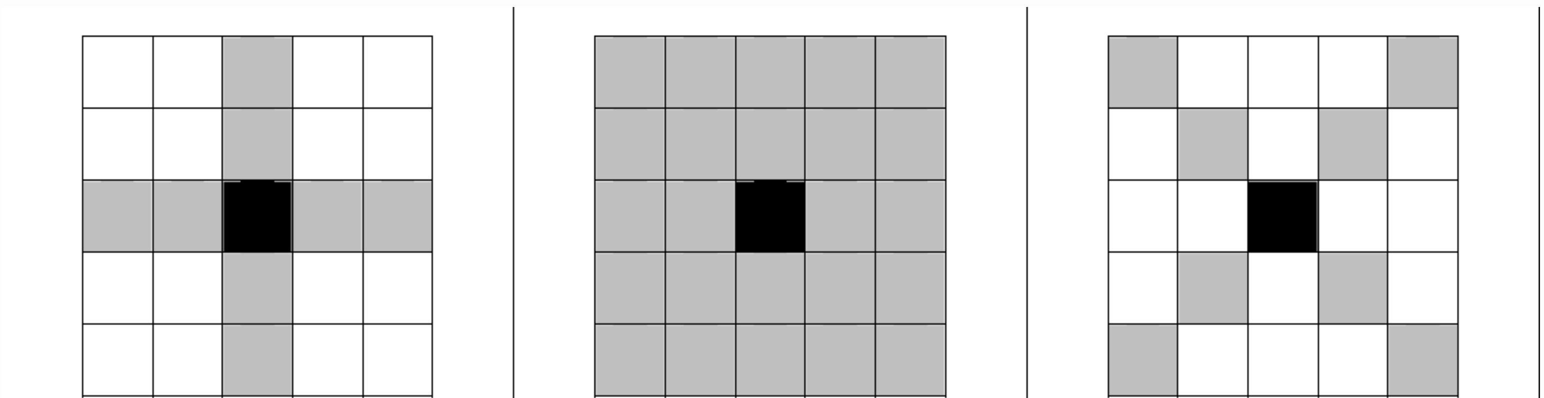
Matriz alfil (bishop) de primer orden

Fuente Gráficos : Baronio, A., Vianco, A. , Rabanal, C. (2012): Una introducción a la econometría espacial

Matrices de contigüidad de órdenes superiores

Dos regiones *i* y *j* serán **contiguas de segundo orden** si ambas están separadas por una tercera región *h* que es contigua de primer orden a ambas.

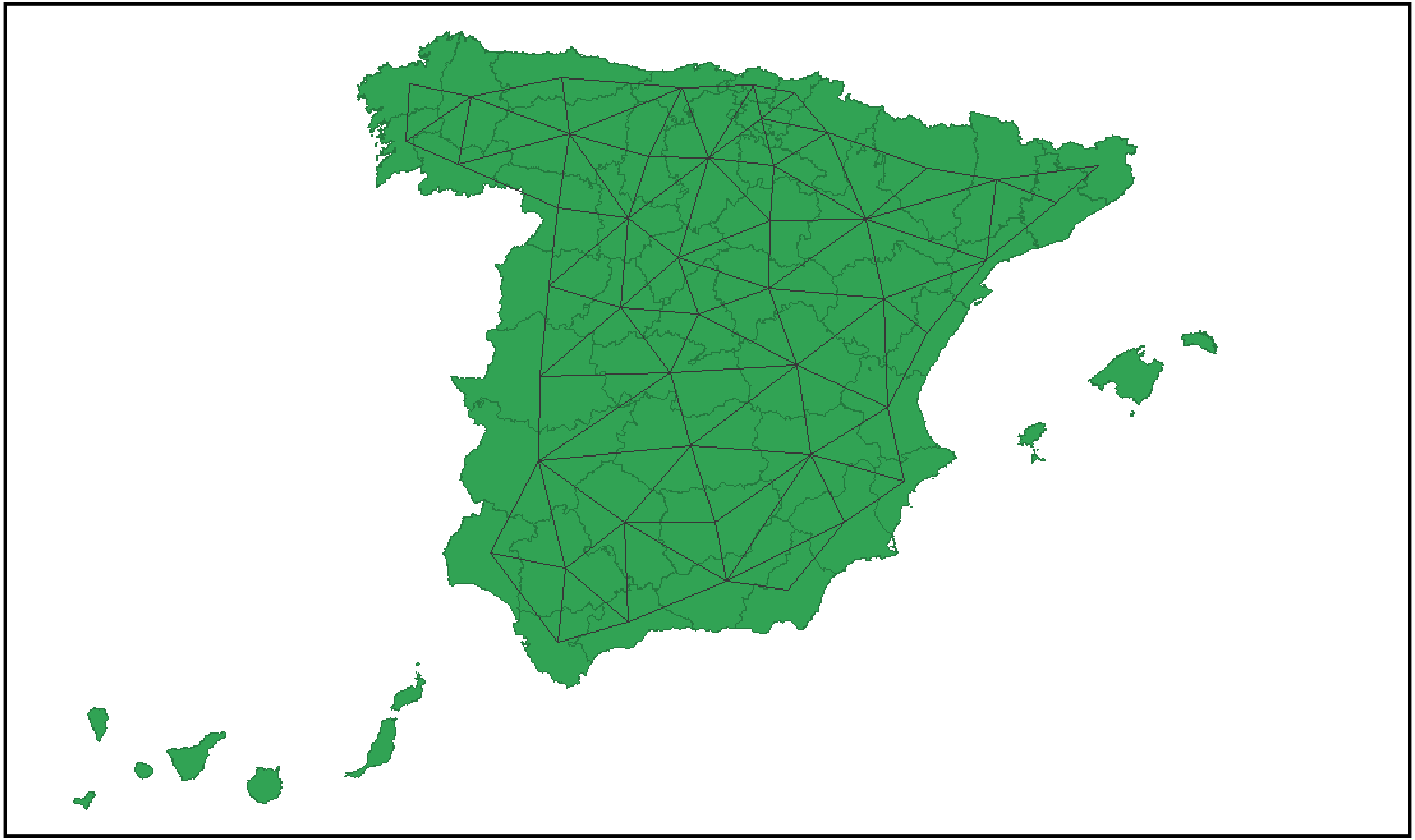
- Aunque la idea es extensible a órdenes superiores, no existen situaciones que requieran adoptar un criterio de vecindad más allá del segundo orden.



Matriz torre (rook) de primer orden | *Matriz reina (queen) de primer orden* | *Matriz alfil (bishop) de primer orden*

Otra alternativa es utilizar los k-vecinos más próximos o los situados a una distancia menor que K

Fuente Gráficos : Baronio, A., Vianco, A. , Rabanal, C. (2012): Una introducción a la econometría espacial



Contigúidad reina matriz de interacciones espaciales

Contigüidad reina para la provincia de Madrid

Matriz W de pesos espaciales: DISTANCIA

Otra alternativa es construir la matriz de pesos por **distancias**:

$$w_{ij} = f(d_{ij}, \theta)$$

con $\frac{\partial w_{ij}}{\partial d_{ij}} < 0$, y siendo d_{ij} una métrica de la distancia entre dos puntos i,j

Por ejemplo: $$w_{ij} = f\left(d_{ij}, \theta\right) = \frac{1}{d_{ij}^{\alpha}}$$ Con d_{ij} la distancia Euclídea

$$w_{ij} = f\left(d_{ij}, \theta\right) = e^{-\beta d_{ij}}$$

Si α=2 (Anselin 1980), la intensidad de la interdependencia entre dos regiones disminuye con la distancia que separa sus respectivos centros.

También se pueden elegir **los k-vecinos más próximos** o los situados a una **distancia menor que K** y estandarizar por filas

Matriz W de pesos espaciales: **NORMALIZACION POR FILAS**

Una vez que tenemos la matriz W con unos y ceros (por contigüidad), o con pesos inversamente proporcionales a la distancia (por distancia) , suele estandarizarse la matriz W por **filas**:

$$w_{i,j}^{s} = \frac{w_{ij}}{\sum_j w_{ij}}$$

De forma que la suma de los pesos asociados a cada área es igual a uno:

$$\sum_j w_{i,j}^{s} = 1 \qquad \sum_i \sum_j w_{i,j}^{s} = n$$

Además, para obtener estimaciones consistentes, la matriz W debe estar delimitada de manera uniforme y en valor absoluto, lo que implica que existe una constante $c<\infty$ tal que:

$$1 \leq i \leq n \sum_{j=1}^{n} |w_{ij}| \leq c \qquad y \qquad 1 \leq j \leq n \sum_{i=1}^{n} |w_{ij}| \leq c$$

Nota: Autocorrelación espacial y autocorrelación temporal

¿En qué se diferencia la autocorrelación espacial de la autocorrelación temporal?

- La autocorrelación temporal es **unidireccional**: el pasado explica el presente.

El instrumento utilizado para representar esta relación es el **operador de retardos L**.

- La autocorrelación espacial es multidireccional: una región puede no solo estar afectada por otra región contigua a ella sino por otras muchas que la rodean, al igual que ella puede influir sobre aquéllas.

El operador retardos espaciales será precisamente la **matriz de pesos espaciales W**.

NO puede tratarse con técnicas econométricas estándar debido a esta **multidireccionalidad**. Requiere de la econometría espacial.

Retardo espacial

La matriz de pesos espaciales **W** sirve para **construir las variables retardadas espacialmente** o los retardos espaciales de una variable dada ***y***. Así dada una matriz de pesos espaciales

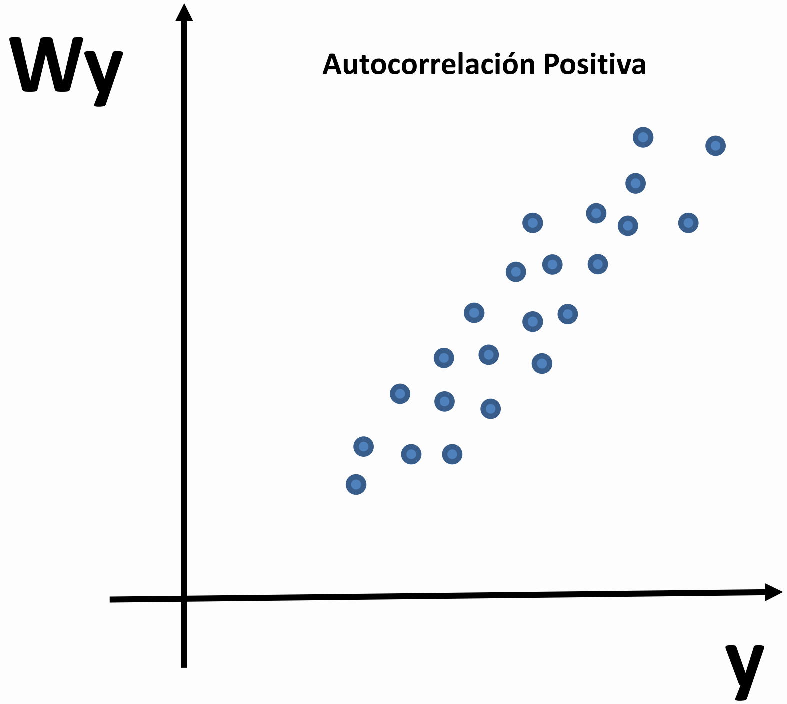
$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix} \qquad w_{i,j}^{s} = \frac{w_{ij}}{\sum_j w_{ij}} \qquad \sum_j w_{i,j}^{s} = 1$$

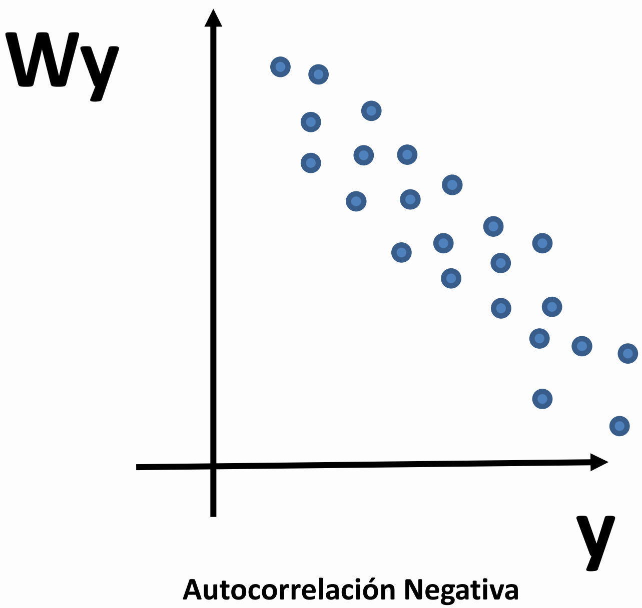
El retardo espacial de una variable dada *y* se construye premultiplicando al vector columna ***y*** (*n*x1)por la matriz pesos espaciales **W** (*n*xn)

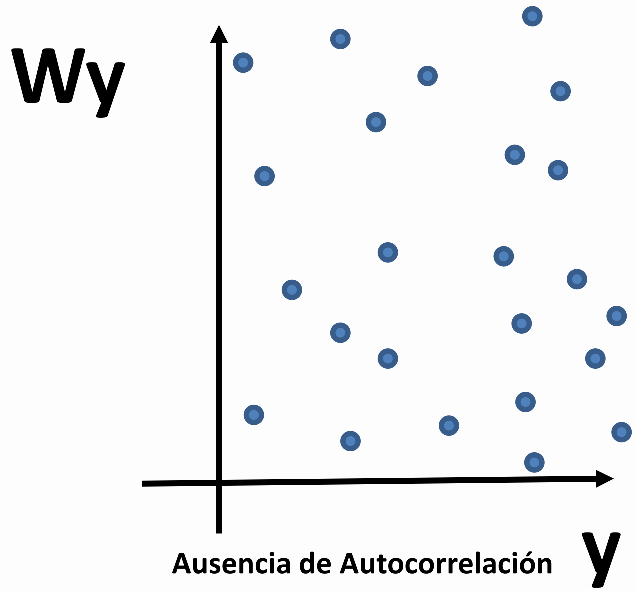
$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \xrightarrow{\text{Retardo espacial}} W \cdot y = \begin{pmatrix} \sum_j w_{1j}^{s} y_j \\ \sum_j w_{2j}^{s} y_j \\ \vdots \\ \sum_j w_{nj}^{s} y_j \end{pmatrix}$$

Cada elemento de este vector ***Wy*** es la media ponderada de los valores de la variable ***y*** en el subgrupo de las observaciones contiguas (vecinas), dado que el peso es cero si no pertenece a ese subgrupo.

Scatterplot de Moran (gráfico de dispersión de Moran)







Estadísticos globales de Autocorrelación Espacial

Una vez que disponemos de la matriz de pesos espaciales, podemos estimar la dependencia o autocorrelación espacial analizando la correlación entre una serie x y su retardo espacial Wx : Medidas de asociación o dependencia espacial **global**:

- El análisis de Autocorrelación Espacial global analiza todas las observaciones de la muestra de forma conjunta para determinar si una variable se encuentra distribuida de forma totalmente aleatoria en el espacio o si, por el contrario, existe un patrón espacial determinado

- Hablamos de **patrón espacial** cuando existe una asociación significativa de valores similares o disímiles entre regiones vecinas.

- Estadísticos globales:
  - ***I*** de Moran,
  - ***C*** de Geary
  - ***G(d)*** de Getis y Ord.

- Permiten contratar la hipótesis nula de "no autocorrelación espacial".

Entre los instrumentos que permiten extraer las características de los datos georreferenciados (observación de una variable asociada a una localización del espacio geográfico) destaca el **índice I de Moran**.

El índice **I de Moran**

$$I = \frac{n}{S_0}\frac{(y-\bar{y})'W(y-\bar{y})}{(y-\bar{y})'(y-\bar{y})} = \frac{n}{S_0}\frac{\sum_i \sum_j w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\sum_i (y_i-\bar{y})^2}$$

Número de zonas en consideración

Covarianza

Varianza

con $S_0 = \sum_i \sum_j w_{i,j}^s$ → Número de relaciones (adyacencias) en el mapa

La I de Moran es una adaptación de una medida de correlación no-espacial al contexto espacial.

La autocorrelación o dependencia espacial detectada puede ser:

- **Negativa**
- **Positiva**
- **Nula (test para contrastar la Hipótesis nula de Ausencia de autocorrelación Espacial) Ho:I=0**

La I de Moran toma valores desde **-1**, correlación espacial perfecta negativa, 0 ausencia de correlación espacial, y **1** correlación espacial perfecta positiva

Modelos de Regresión: Análisis de la asociación entre una variable Objetivo (Y) y una o varias variables explicativas (X)

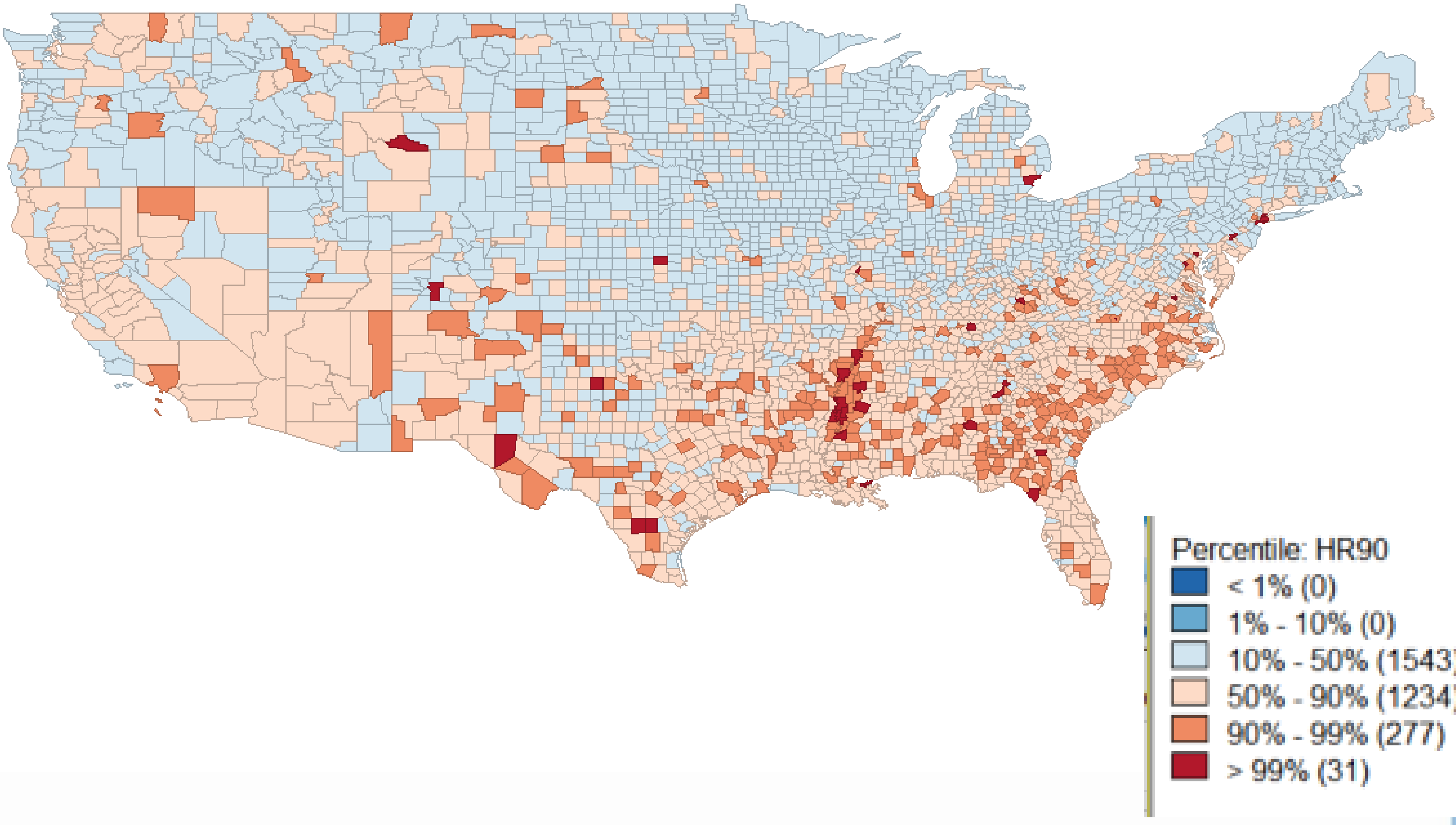
Cuando el análisis exploratorio de la Variable Y muestra la existencia de dependencia ESPACIAL es necesario que el modelo explique dicha dependencia espacial. Toda la dependencia espacial de la variable Y que no sea explicada por el modelo quedará reflejada en los RESIDUOS

La diagnosis del modelo:
¿Son los residuos un Ruido Blanco? ¿se distribuyen con media cero, varianza constante y distribuidos aleatoriamente por el espacio?, es decir, sin dependencia espacial.

Cuando el modelo no es capaz de explicar la dependencia espacial de la variable Y (los residuos presentan autocorrelación espacial, será necesario incluir de alguna forma dicha dependencia espacial de forma explícita en el modelo econométrico hasta que los residuos sea ruido blanco (se distribuyan de manera aleatoria por todo el territorio).

Ejemplo: Vamos a construir un modelo para intentar explicar la tasa de Homicidios, que sabemos que se encuentra autocorrelacionada espacialmente

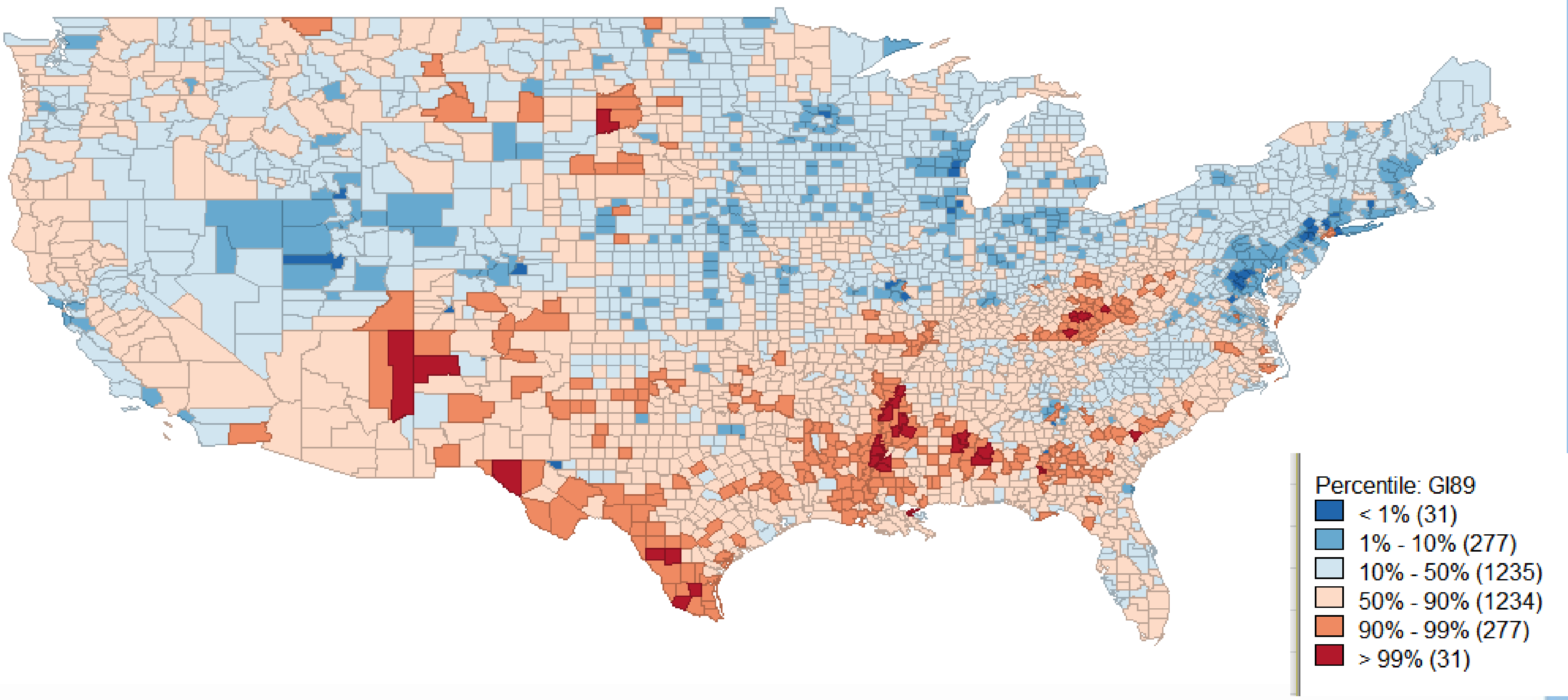
Homicide rate per 100,000 (1990)



Como Variable explicativa de los homicidios vamos a utilizar el índice de GINI

$$Homicidios_j = b0 + b1*Gini_j + u_j$$

Gini index of family income inequality in 1990



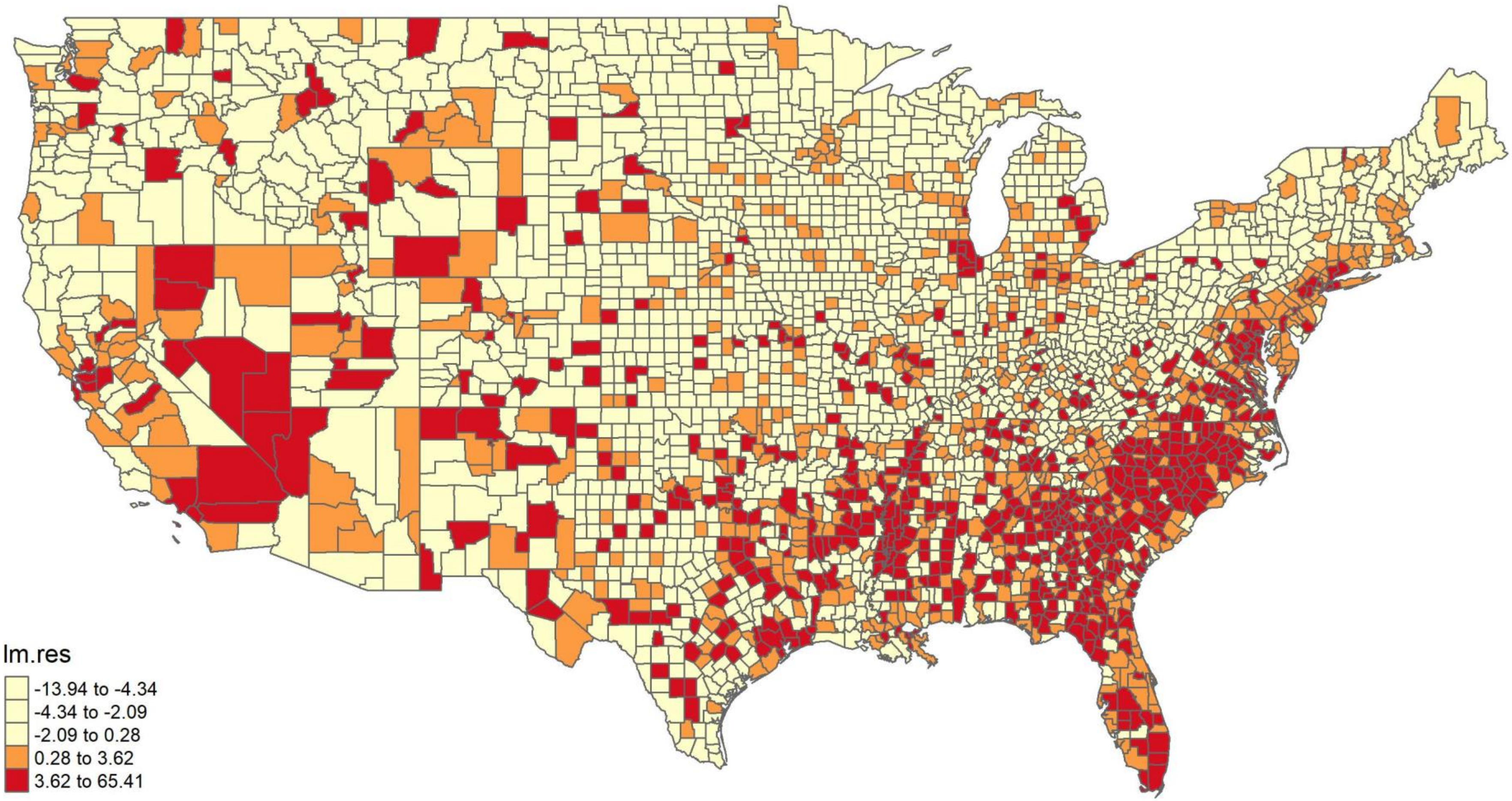
Como Variable explicativa de los homicidios vamos a utilizar el índice de GINI

$$Homicidios_j = b0 + b1*Gini_j + u_j$$

```
Call:
lm(formula = HR90 ~ GI89, data = NAT@data)

Residuals:
    Min      1Q  Median      3Q     Max
-13.936  -3.698  -0.936   2.608  65.412

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -26.798      1.118  -23.97   <2e-16 ***
GI89          88.017      2.970   29.64   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.86 on 3083 degrees of freedom
Multiple R-squared:  0.2217,    Adjusted R-squared:  0.2215
F-statistic: 878.3 on 1 and 3083 DF,  p-value: < 2.2e-16
```

La diagnosis del modelo: ¿Explica el índice de Gini la tasa de Homicidios?¿Y explica también su estructura de autocorrelación espacial?¿son los residuos del modelo un ruido blanco? ¿tienen los residuos un patrón espacial?: I de moran sobre los residuos de un modelo de regresión lineal

Residuos del modelo de regresión lineal



Test de I de Moran sobre los Residuos del modelo de regresión lineal

```
        Global Moran I for regression residuals

data:
model: lm(formula = HR90 ~ GI89, data = NAT@data)
weights: W

Moran I statistic standard deviate = 27.121, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:
Observed Moran I      Expectation         Variance
    0.2911628429    -0.0005320636     0.0001156769
```

Autocorrelación Espacial en el modelo de regresión

Una vez detectada la presencia de Autocorrelación Espacial a nivel **univariante**, comprobamos si esta **dependencia espacial está presente en el modelo de regresión**:

- **Sustantiva**: Si se omite erróneamente el **retardo espacial de la variable endógena** (Wy) y/o **exógenas** (WX), se traslada al término de error que pasaría a estar autocorrelacionado espacialmente.

Solución: incluir en el modelo un retardo espacial de la variable endógena o exógena correlacionada espacialmente.

- **Residual**: omisión de variables no relevantes que están correlacionadas espacialmente o la existencia de errores de medida.

Solución: Modelización explícita del esquema de dependencia espacial en el término de error.

Modelos de regresión espacial

- 1) **Modelo con retardo espacial** en la variable dependiente (SLM/SLAG) o **modelo espacial autorregresivo** (**SAR**):

$$\boldsymbol{y} = \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{u}$$

$$\boldsymbol{u} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I})$$

a) Parámetro autorregresivo (ρ) que recoge la intensidad de las interdependencias entre las observaciones muestrales.

b) Las estimaciones MCO serán <u>sesgadas e inconsistentes</u> ante la presencia de un retardo espacial en la variable endógena, aun cuando los errores no estén correlacionados espacialmente. Hay que utilizar MC2E (o métodos GM) utilizando como instrumentos los retardos espaciales del resto de variables Explicativas. Además deben estimarse siempre consistentemente la verdadera matriz de varianzas y covarianzas de los estimadores

- 2) Modelo espacial de Durbin (**SDM**) o modelo factor común:

$$\boldsymbol{y} = \rho\mathbf{W}\boldsymbol{y} + \mathbf{X\beta} + \mathbf{WX}\boldsymbol{\gamma} + \boldsymbol{u}$$
$$\boldsymbol{u} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I})$$

- 3) Modelo espacial regresivo-cruzado (cross-reggresive) (**SCM**) o modelo con retardo espacial de X (**SLX**):

$$\boldsymbol{y} = \mathbf{X\beta} + \mathbf{WX}\boldsymbol{\gamma} + \boldsymbol{u}$$

$$\boldsymbol{u} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I})$$

Modelos de regresión espacial

4) Modelo de regresión lineal con error autorregresivo espacial o **modelo de error espacial** (**SEM**):

$$\boldsymbol{y} = \mathbf{X\beta} + \boldsymbol{u}$$

$$\boldsymbol{u} = \boldsymbol{\lambda}\mathbf{W}\boldsymbol{u} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I})$$

a) Parámetro autorregresivo (λ) que recoge la intensidad de las interdependencias.
b) Ante la presencia de errores correlacionados espacialmente, las estimaciones MCO seguirán siendo insesgadas pero ineficientes porque la <u>matriz de varianzas-covarianzas será no esférica</u>. Debemos estimarla consistentemente.

5) Modelo espacial autorregresivo combinado (**SAC**) (Kelijian-Prucha model):

$$\boldsymbol{y} = \boldsymbol{\rho}\boldsymbol{W}_1\boldsymbol{y} + \boldsymbol{X\beta} + \boldsymbol{u}$$
$$\boldsymbol{u} = \boldsymbol{\lambda}\boldsymbol{W}_2\boldsymbol{u} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I})$$

Podemos suponer:

$$\mathbf{W_1} = \mathbf{W_2} = \mathbf{W}$$

6) Modelo de error espacial de Durbin (**SDEM**):

$$\boldsymbol{y} = \mathbf{X\beta} + \boldsymbol{W}_1\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{u}$$

$$\boldsymbol{u} = \boldsymbol{\lambda}\boldsymbol{W}_2\boldsymbol{u} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I})$$

Modelos de regresión espacial

7) General Nesting Spatial Model (**GNSM**) (Manski model):

Podemos suponer:

$$\mathbf{W_1} = \mathbf{W_2} = \mathbf{W}$$

$$\boldsymbol{y} = \rho \boldsymbol{W_1 y} + \mathbf{X\beta} + \boldsymbol{W_2} \mathbf{X} \boldsymbol{\gamma} + \boldsymbol{u}$$

$$\boldsymbol{u} = \boldsymbol{\lambda} \mathbf{W} \boldsymbol{u} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma}^2 \mathbf{I})$$

a) A partir del modelo **GNSM** podemos obtener los anteriores imponiendo restricciones sobre los parámetros: λ=0, ρ=0, y θ=0.

8) Modelo espacial autorregresivo y media móvil (**SARMA**):

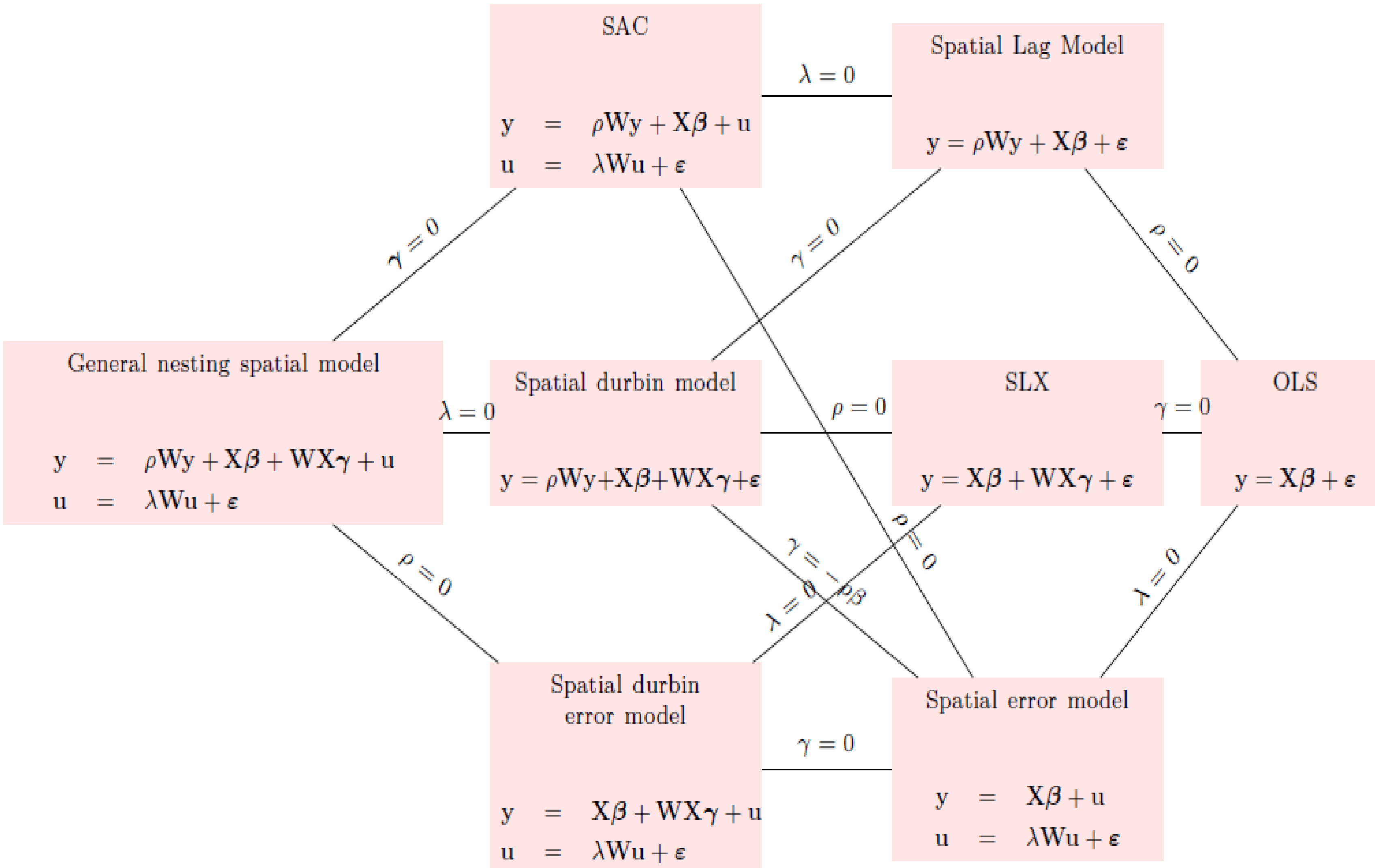
Podemos suponer:

$$\mathbf{W_1} = \mathbf{W_2} = \mathbf{W}$$

$$\boldsymbol{y} = \rho \boldsymbol{W_1 y} + \mathbf{X\beta} + \boldsymbol{u}$$

$$\boldsymbol{u} = \boldsymbol{\varepsilon} - \boldsymbol{\theta} \boldsymbol{W_2} \boldsymbol{\varepsilon} + \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma}^2 \mathbf{I})$$

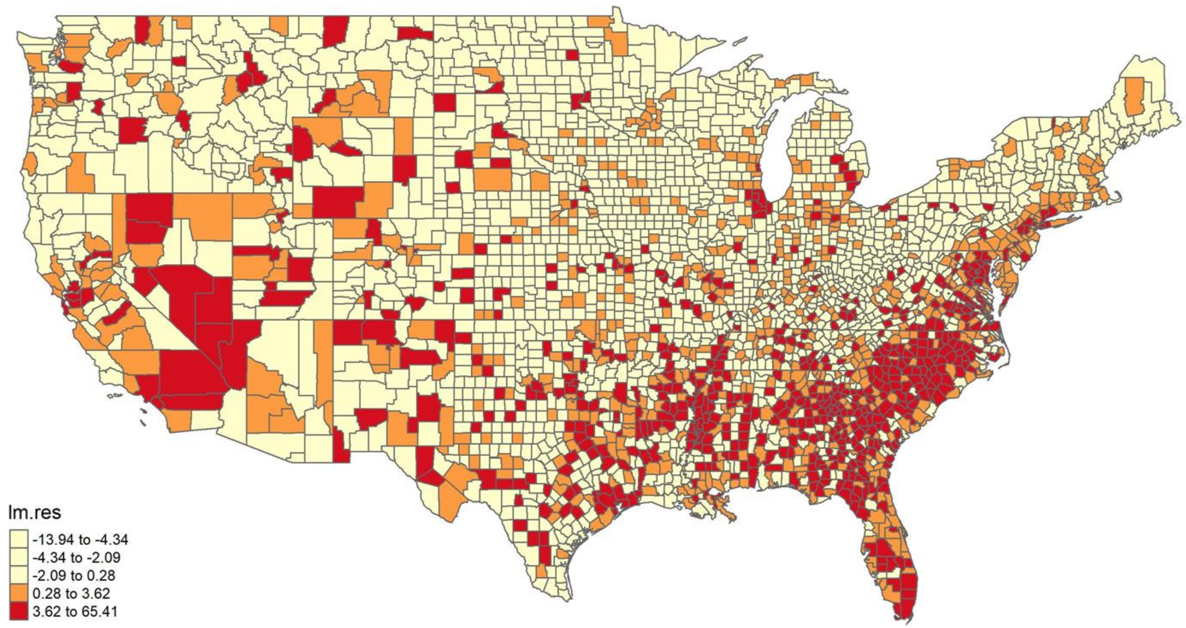
a) La diferencia entre el modelo **SAC** y el modelo **SARMA** está en la estructura del término de error: autorregresiva vs. media móvil.



SAC
y = ρWy + Xβ + u
u = λWu + ε
Spatial Lag Model
y = ρWy + Xβ + ε
λ = 0
γ = 0
γ = 0
ρ = 0
General nesting spatial model
y = ρWy + Xβ + WXγ + u
u = λWu + ε
λ = 0
Spatial durbin model
y = ρWy+Xβ+WXγ+ε
ρ = 0
SLX
y = Xβ + WXγ + ε
γ = 0
OLS
y = Xβ + ε
ρ = 0
γ = −ρβ
λ = 0
ρ = 0
λ = 0
Spatial durbin
error model
y = Xβ + WXγ + u
u = λWu + ε
γ = 0
Spatial error model
y = Xβ + u
u = λWu + ε

Residuos del Modelo Lineal

```
        Global Moran I for regression residuals

data:
model: lm(formula = HR90 ~ GI89, data = NAT@data)
weights: W

Moran I statistic standard deviate = 27.121, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:
Observed Moran I      Expectation         Variance
    0.2911628429    -0.0005320636     0.0001156769
```



Residuos de modelo espacial

```
        Moran I test under randomisation

data:  NAT$m7.res
weights: W

Moran I statistic standard deviate = -1.4154, p-value = 0.9215
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
    -0.0155124064     -0.0003242542      0.0001151489
```