

# *Geographic business intelligence*

Francesc Puigvert i Pell

PID\_00210845



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

# Índice

<b>Introducción</b> .....	5
<b>1. Importancia de los datos geográficos en los sistemas de <i>business intelligence</i></b> .....	7
1.1. Cuadrante mágico de Gartner .....	11
<b>2. Sistemas de información geográfica</b> .....	14
<b>3. Los SIG en los procesos de analítica de negocio</b> .....	16
3.1. Público objetivo y beneficios .....	16
<b>4. Los datos geográficos</b> .....	18
4.1. Componente espacial .....	18
4.1.1. Localización .....	18
4.1.2. Modelos de representación .....	22
4.1.3. Atributos espaciales .....	25
4.1.4. Relaciones espaciales .....	25
4.2. Componente temático .....	28
4.2.1. Clasificación en función del modelo .....	28
4.2.2. Clasificación en función de la naturaleza .....	30
4.3. Componente temporal .....	33
<b>5. Comunicación cartográfica</b> .....	35
5.1. Métodos de representación cartográfica .....	36
5.1.1. Representación simple .....	36
5.1.2. Clasificación por categorías .....	37
5.2. Técnicas de representación en mapas .....	42
5.2.1. Agrupaciones o <i>clustering</i> .....	42
5.2.2. Mapa de calor ( <i>heatmap</i> ) .....	43
5.2.3. Mapas de flujos .....	44
5.2.4. Colores graduados o coropletas .....	45
5.2.5. Símbolos graduados .....	46
5.2.6. Puntos de densidad .....	47
<b>6. Análisis de datos</b> .....	48
6.1. Operaciones de análisis básicas .....	48
6.2. Operaciones de <i>geomarketing</i> .....	51
6.2.1. Diseño del territorio .....	51
6.2.2. Planificación de mercado .....	56
<b>7. Fuentes de datos externas</b> .....	59

7.1. Normalización para proveer de geolocalización .....	59
7.2. Servicios de callejero .....	60
7.3. Variables de mercado .....	62
<b>8. Soluciones de mercado.....</b>	<b>64</b>
8.1. Soluciones SIG .....	64
8.2. Soluciones BI .....	66
<b>Resumen.....</b>	<b>67</b>
<b>Glosario.....</b>	<b>69</b>
<b>Bibliografía.....</b>	<b>71</b>

## Introducción

Hoy en día es habitual que al comprar productos en determinadas tiendas el cajero nos pregunte: “¿código postal?”. La razón de esta pregunta es que, seguramente, la empresa realiza estudios de mercado con los datos de los clientes. Los objetivos pueden ser variados, como por ejemplo lanzar campañas más efectivas por segmento de población, identificar el nicho para la ubicación de nuevas tiendas o identificar tipos de cliente cruzando el código postal con variables socioeconómicas tipo, como puede ser el nivel adquisitivo medio de los habitantes de un código postal determinado.

En estos materiales asumiremos que el lector ya conoce los procesos relacionados con la actividad de inteligencia de negocio y que ya sabe las aportaciones que las soluciones de *business intelligence* pueden proporcionar. El material se enfoca a las situaciones donde la actividad de inteligencia de negocio utiliza información geográfica. En estos casos los sistemas de información geográfica son un elemento que hay que tener en cuenta. A continuación presentamos un conjunto de preguntas que un responsable de cuentas puede haberse hecho en su actividad profesional.

- ¿Dónde puedo abrir mi nuevo negocio?
- ¿Dónde elegir la mejor ubicación sobre el territorio?
- ¿Dónde están localizados mis recursos?
- ¿Dónde están mis clientes potenciales?
- ¿Desde qué zona domino el mercado?
- ¿Dónde está localizada mi competencia?
- ¿Existe alguna tipología de clientes común o siguen patrones distintos?
- ¿Desde qué zona existe alguna correlación entre los patrones?
- ¿Tras que patrón existe alguna correlación entre las distintas zonas?
- ¿Existen zonas similares que ofrecen la misma productividad? ¿Y productividad distinta?
- ¿Qué zona me ofrece un mejor rendimiento?

- ¿Tengo algún patrón que explique el comportamiento de mi empresa y lo puedo reproducir en una zona similar?
- ¿Puedo replicar mi zona de éxito en sitios con clientes similares?
- ¿Cuál es la mejor manera de distribuir la fuerza de ventas sobre el territorio?

El material está enfocado a que el lector entienda cómo funciona la información geográfica y cómo integrarla en los sistemas de inteligencia de negocio para dar respuesta a preguntas similares a las anteriores.

# 1. Importancia de los datos geográficos en los sistemas de business intelligence

Todos los elementos y eventos físicos tienen una ubicación determinada. Esa ubicación, si puede almacenarse y procesarse, puede resultar muy útil, no solo desde un punto de vista operativo sino también en los procesos analíticos. De hecho, la información geográfica es relevante desde dos puntos de vista diferenciados: la representación de información (para representar el lugar de ocurrencia de los hechos y agentes del mundo real) y la visualización de información (para mostrar un conjunto de datos teniendo en cuenta sus características geográficas). A continuación mostramos, mediante un ejemplo, la importancia de los datos geográficos en los sistemas de inteligencia de negocio (BI de aquí en adelante) y analizamos los informes de algunas consultoras de prestigio para dar fuerza a esta afirmación.

Supongamos que un analista revisa una lista de ventas por clientes en el área de Barcelona similar a la imagen inferior:

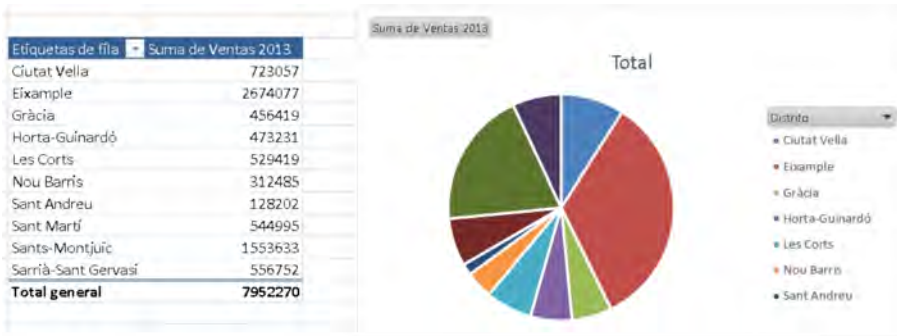
Ventas por cliente en Barcelona

Activo	Comercial	Tipo	Ventas 2011	Ventas 2012	Ventas 2013	Calle	Ciudad	CP	Código postal	Distrito	Cuarto	Barrio	
1	S	Comercial C	11.897,00 €	11.630,00 €	9.998,00 €	PASSEIG GRÀCIA 15	BARCELONA	80070	238	Eixample	7	Eixample Dret	2,169
2	S	Comercial A	20.840,00 €	18.884,00 €	17.094,00 €	AVINGUDA DIAGONAL 537	BARCELONA	80290	115	Les Corts	19	les Corts	2,133
3	S	Comercial A	11.503,00 €	12.315,00 €	12.682,00 €	CARRER MUNTANER 65	BARCELONA	80110	239	Eixample	8	Antiga Eixample Esquerra	2,155
4	S	Comercial A	16.262,00 €	18.172,00 €	16.951,00 €	AVINGUDA DIAGONAL 639	BARCELONA	80280	19	Les Corts	20	la Maternitat i Sant Ramon	2,121
5	S	Comercial C	22.603,00 €	22.831,00 €	23.507,00 €	AVINGUDA GAUDÍ 58	BARCELONA	80250	266	Eixample	6	la Sagrada Família	2,174
6	S	Comercial C	14.738,00 €	13.990,00 €	15.126,00 €	AVINGUDA GAUDÍ 58	BARCELONA	80250	297	Eixample	6	la Sagrada Família	2,174
7	S	Comercial C	16.392,00 €	16.316,00 €	15.137,00 €	AVINGUDA GAUDÍ 58	BARCELONA	80250	392	Eixample	6	la Sagrada Família	2,174
8	S	Comercial C	15.106,00 €	16.227,00 €	17.548,00 €	AVINGUDA MERIDIANA	BARCELONA	80260	68	Sant Martí	64	el Camp Arpa del Clot	2
9	S	Comercial C	23.754,00 €	23.594,00 €	23.325,00 €	AVINGUDA ROMA 51	BARCELONA	80290	261	Eixample	9	Nou Eixample Esquerra	2,14
10	S	Comercial A	22.053,00 €	20.630,00 €	19.937,00 €	AVINGUDA ROMA 51	BARCELONA	80290	21	Eixample	9	Nou Eixample Esquerra	2,14
11	S	Comercial A	10.740,00 €	12.591,00 €	11.960,00 €	CARRER PROVENÇA 64	BARCELONA	80290	6	Eixample	9	Nou Eixample Esquerra	2,143
12	S	Comercial A	19.223,00 €	19.517,00 €	20.809,00 €	AVINGUDA SARRIÀ 55	BARCELONA	80290	243	Les Corts	19	les Corts	2,142
13	S	Comercial B	12.403,00 €	13.593,00 €	14.598,00 €	JARDINS DEL PORT ROMÀ 154	BARCELONA	80380	7	Sants-Montjuïc	13	la Marina de Port	2,143
14	S	Comercial C	10.895,00 €	11.514,00 €	13.057,00 €	CARRER ALMOGÀVERS 110	BARCELONA	80180	245	Sant Martí	66	el Parc i la Llacuna del Poblenou	2,185
15	S	Comercial A	13.638,00 €	14.493,00 €	13.857,00 €	CARRER TORNIS 28	BARCELONA	80280	8	Sants-Montjuïc	17	Sants - Badal	2,124
16	S	Comercial A	19.245,00 €	20.505,00 €	20.452,00 €	CARRER ALCOLEA 19	BARCELONA	80140	435	Sants-Montjuïc	18	Sants	2,136
17	S	Comercial C	8.839,00 €	8.679,00 €	10.143,00 €	CARRER VALÈNCIA 423	BARCELONA	80130	9	Eixample	6	la Sagrada Família	2,175
18	S	Comercial A	10.605,00 €	11.576,00 €	13.155,00 €	CARRER ARIBAU 162	BARCELONA	80360	247	Eixample	8	Antiga Eixample Esquerra	2,157
19	S	Comercial A	10.623,00 €	10.287,00 €	8.462,00 €	CARRER CASANOVA 45	BARCELONA	80110	10	Eixample	8	Antiga Eixample Esquerra	2,160
20	S	Comercial B	9.970,00 €	10.721,00 €	11.490,00 €	CARRER REC COMTAL 15	BARCELONA	80030	11	Ciutat Vella	4	Sant Pere Santa Caterina i la Ribera	2,175
21	S	Comercial B	20.294,00 €	21.360,00 €	21.021,00 €	CARRER ALLADA VERMELL 3	BARCELONA	80030	359	Ciutat Vella	4	Sant Pere Santa Caterina i la Ribera	2,181
22	N	Comercial D	14.697,00 €	13.189,00 €	12.093,00 €	PASSEIG PEIRA 3	BARCELONA	80310	12	Nou Barris	46	el Turó de la Peira	2,116
23	S	Comercial C	20.972,00 €	22.041,00 €	20.771,00 €	CARRER ALMOGÀVERS 148	BARCELONA	80180	391	Sant Martí	66	el Parc i la Llacuna del Poblenou	2,15
24	S	Comercial B	18.701,00 €	20.458,00 €	20.106,00 €	CARRER ALTS FORNS	BARCELONA	80380	241	Sants-Montjuïc	13	la Marina de Port	2,1
25	S	Comercial B	20.766,00 €	19.360,00 €	19.644,00 €	CARRER AMPOSTA 26	BARCELONA	80040	302	Sants-Montjuïc	14	la Font de la Guatlúa	2,148
26	S	Comercial B	17.965,00 €	19.086,00 €	17.472,00 €	CARRER ARENES DELS CANVIS 1	BARCELONA	80030	317	Ciutat Vella	4	Sant Pere Santa Caterina i la Ribera	2,181
27	S	Comercial A	14.667,00 €	14.130,00 €	14.773,00 €	CARRER MOIÀ 1	BARCELONA	80060	251	Sarrià-Sant Gervasi	26	Sant Gervasi - Galvany	2,152
28	S	Comercial D	19.295,00 €	18.668,00 €	17.678,00 €	CARRER ARGIMON 24	BARCELONA	80320	199	Barcelona	37	el Carmel	2,155
29	S	Comercial A	21.204,00 €	21.075,00 €	21.484,00 €	CARRER ARIBAU 46	BARCELONA	80110	130	Eixample	8	Antiga Eixample Esquerra	2,160
30	S	Comercial A	19.644,00 €	20.091,00 €	21.945,00 €	CARRER ARIZALA 49	BARCELONA	80280	409	Les Corts	20	la Maternitat i Sant Ramon	2,122
31	S	Comercial A	17.273,00 €	18.842,00 €	16.989,00 €	CARRER ARIZALA 58	BARCELONA	80280	116	Les Corts	20	la Maternitat i Sant Ramon	2,122
32	S	Comercial B	17.855,00 €	19.742,00 €	18.589,00 €	CARRER ATÀULF 5	BARCELONA	80020	292	Ciutat Vella	2	el Barri Gòtic	2,178
33	S	Comercial B	18.663,00 €	18.955,00 €	17.235,00 €	CARRER ATLÀNTIDA 67	BARCELONA	80030	326	Ciutat Vella	3	la Barceloneta	2,190
34	S	Comercial A	12.922,00 €	11.269,00 €	11.310,00 €	CARRER MALLORCA 167	BARCELONA	80360	256	Eixample	8	Antiga Eixample Esquerra	2,155
35	S	Comercial B	14.988,00 €	15.658,00 €	14.141,00 €	CARRER GRASES	BARCELONA	80040	257	Sants-Montjuïc	11	el Poble Sec	2
36	S	Comercial A	15.803,00 €	14.572,00 €	14.282,00 €	AVINGUDA DIAGONAL 561	BARCELONA	80290	258	Les Corts	19	les Corts	2,13
37	S	Comercial A	19.876,00 €	19.176,00 €	19.106,00 €	CARRER AVENIR 56	BARCELONA	80210	306	Sarrià-Sant Gervasi	26	Sant Gervasi - Galvany	2,145
38	S	Comercial A	21.881,00 €	21.796,00 €	23.293,00 €	CARRER AVENIR 74	BARCELONA	80210	373	Sarrià-Sant Gervasi	26	Sant Gervasi - Galvany	2,144

Fuente: Elaboración propia (emulado).

El analista intenta analizar la distribución de las ventas por distrito. Para ello agrupa los valores y genera un gráfico de tarta que representa los distritos y en el que se muestra el volumen de ventas en el ejercicio del 2013.

## Ventas 2013 por distrito en Barcelona



Fuente: Elaboración propia (emulado).

En la imagen superior, se puede ver que el volumen máximo de ventas se da en el barrio de L'Eixample y en Sants-Monjuic. En este caso, aunque el analista no sea de Barcelona y no conozca la distribución de la ciudad, el gráfico podría serle útil.

Sin embargo, si intenta realizar el mismo ejercicio utilizando el volumen de ventas por barrio, la lectura gráfica del resultado es casi imposible, dado el gran número de barrios existentes, eso sin contar con que el analista pierde la vista de la distribución de los datos en la ciudad de Barcelona porque cada venta, cada cliente, está en algún lugar. Por tanto, el analista perdería el valor de situación, o como se denomina en la comunidad el valor "dónde" del dato.

## Ventas 2013 por barrio en Barcelona



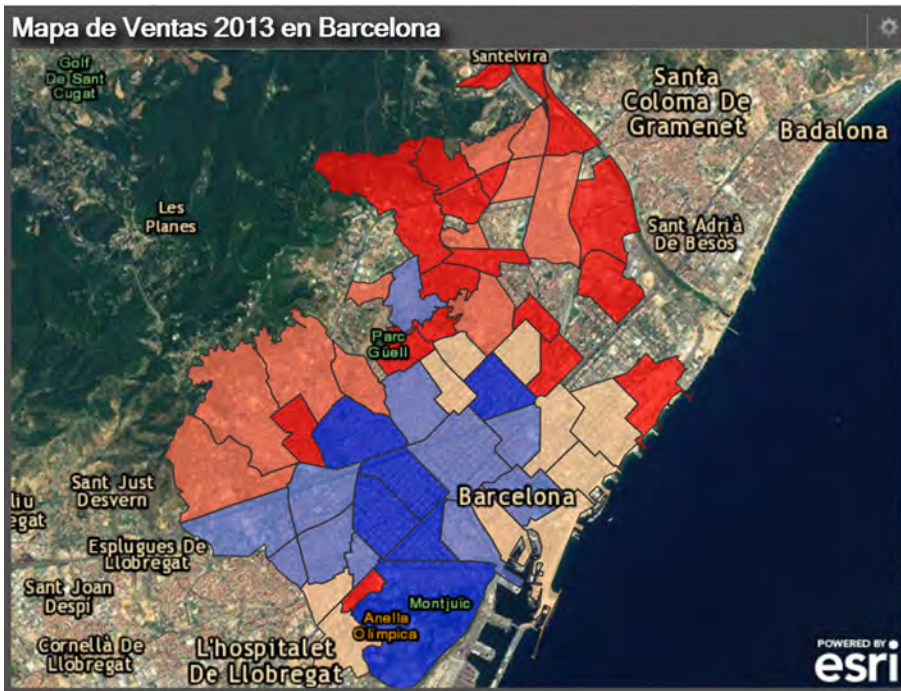
Fuente: Elaboración propia (emulado).

Utilizando un mapa como el de la imagen inferior la lectura es notablemente más fácil: las zonas rojas (en la periferia) son las zonas con menos importe en ventas, mientras que los barrios centrales (en azul) indican un mayor número de ventas.

El mapa no solo es capaz de representar la información de manera efectiva, también es capaz de analizar la existencia de una correlación de ventas entre barrios: las zonas con menos ventas están en la periferia y la zona centro es la que ofrece mayor volumen de ventas.



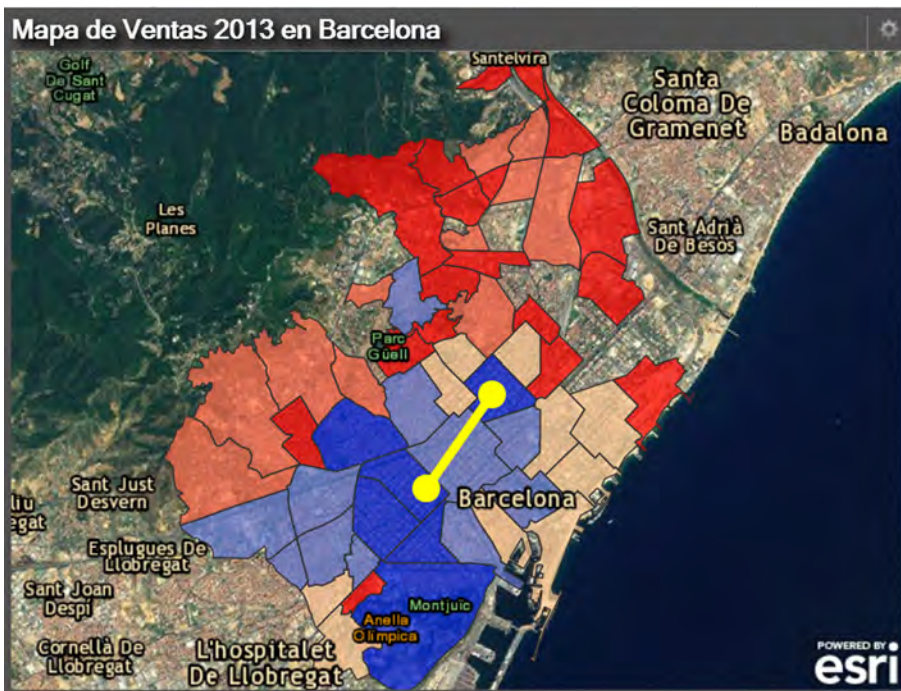
Ventas 2013 por barrios en Barcelona



Fuente: Elaboración propia. Mapa base Esri.

En la figura anterior se observa un salto en la correlación de las ventas en la zona del centro de Barcelona. La línea amarilla de la siguiente figura indica el salto entre zonas. Este comportamiento anómalo en los datos sería un caso de estudio típico de análisis geográfico.

Zona de análisis. Ventas 2013 por barrio en Barcelona

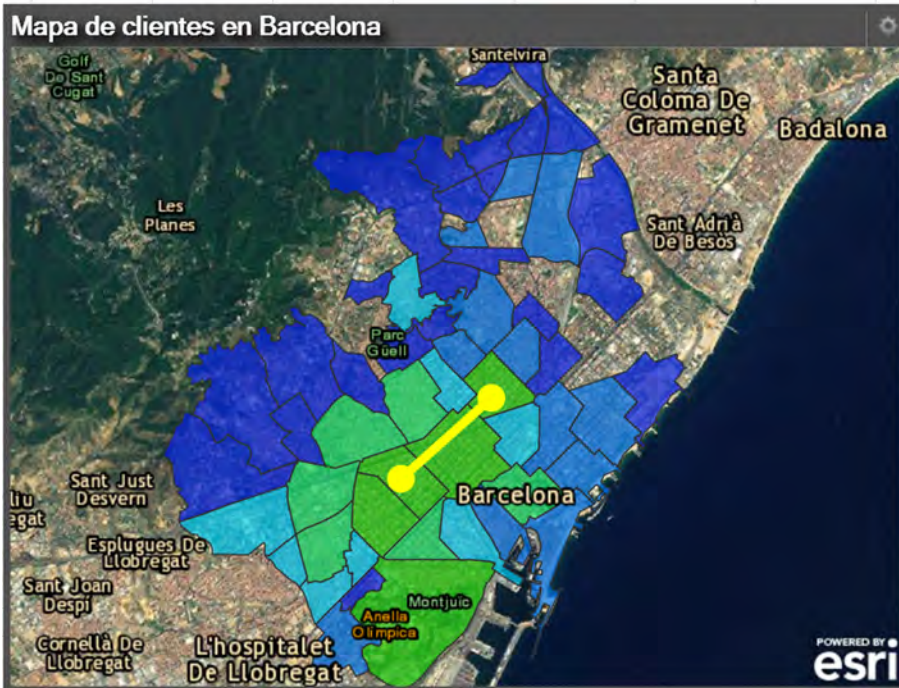


Fuente: Elaboración propia. Mapa base Esri.

Si se representa la información por conteo de clientes, de tal modo que los colores azules correspondan a los barrios con menos clientes y los colores verdes a las zonas con un número de clientes mayor, se observa cómo el número

de clientes en la zona central sigue una disposición regular (línea amarilla), lo que indica que esta zona ofrece un rendimiento de las ventas por debajo de su supuesto potencial.

Número de clientes por barrio en Barcelona



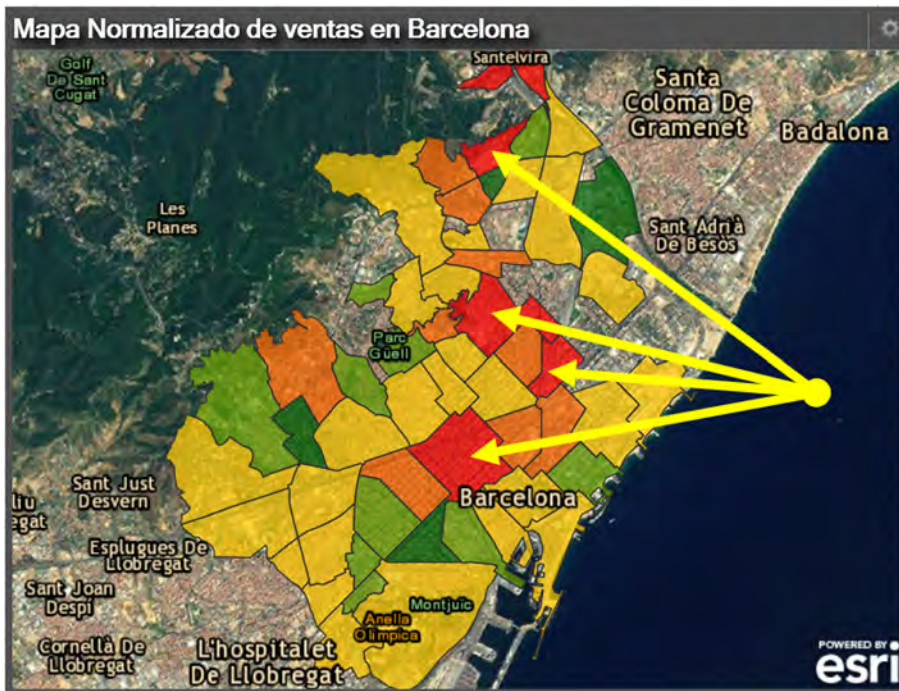
Fuente: Elaboración propia. Mapa base Esri.

La siguiente figura ilustra mejor la problemática en el estudio de la anomalía detectada. En este caso se ha dividido el total de las ventas realizadas en el 2013 entre el número de clientes totales que hay por barrio. Este proceso de estandarización (ventas/clientes) se denomina normalizar los datos. Cuando se pretende englobar los valores de distintos campos en una única representación, es necesario homogeneizar su estructura.

En la siguiente imagen se observan, de color verde, las zonas que ofrecen un mayor importe en ventas en función del número de clientes, mientras que las zonas rojas indican dónde es menor la productividad. Las flechas amarillas (sobrepuestas en el mapa) resaltan las zonas en las que el rendimiento de las ventas en relación con el número de clientes es menor.



Mapa normalizado de ventas por clientes en los barrios de Barcelona



Fuente: Elaboración propia. Mapa base Esri.

Los datos geográficos son de vital importancia para los sistemas de BI, siempre y cuando algunas de las preguntas que se quieran resolver tengan en cuenta información geográfica. Hoy en día, la importancia y expansión de los dispositivos móviles y de las aplicaciones basadas en localización, la democratización de los sistemas de información geográfica (SIG en adelante) y la madurez de los sistemas de BI han provocado que el uso de información geográfica en los sistemas BI sea uno de los principales retos que superar. De hecho, los recientes informes de las consultoras en el sector confirman este argumento.

### 1.1. Cuadrante mágico de Gartner

Cada año la consultora tecnológica Gartner<sup>1</sup> publica distintos estudios comparativos que dan una foto del estado del arte respecto a la tecnología analizada. Del estudio se extrae una representación gráfica que ilustra la situación del momento. Esta representación es conocida como Cuadrante mágico de Gartner.

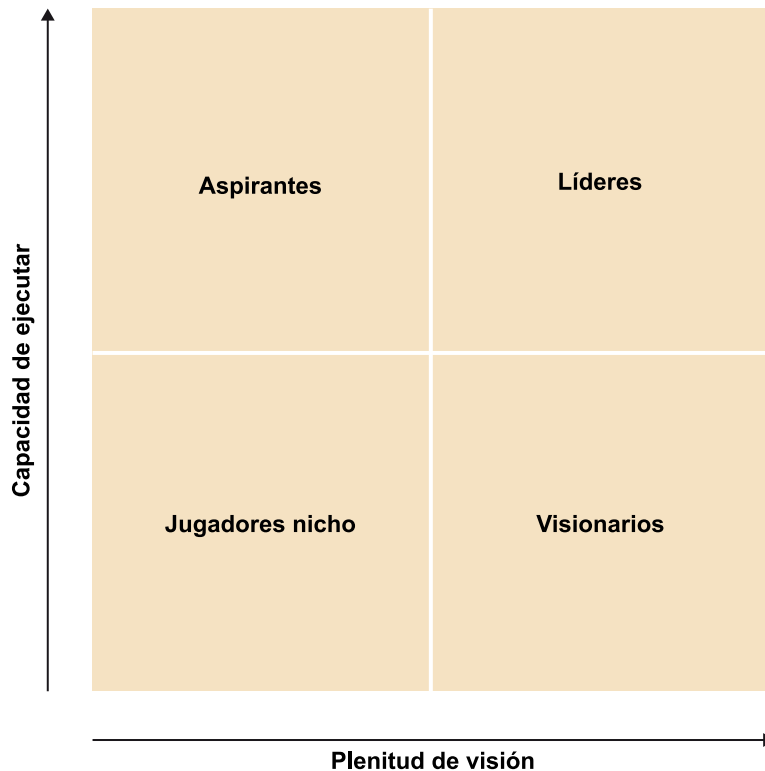
<sup>(1)</sup>Gartner Inc. es una prestigiosa empresa consultora y de investigación sobre tecnologías de la información: <http://www.gartner.com/>

En el cuadrante se identifican cuatro grupos, categorizados en:

- Líderes (*leaders*). Forman parte de este grupo aquellos productos que ejecutan correctamente las funcionalidades del producto actual y están bien posicionados respecto a las tendencias futuras.
- Visionarios (*visionaries*). Productos innovadores pero con una base de clientes limitada.

- Jugadores nicho (*niche players*). Enfocados a un segmento de mercado especializado, pero sin ofrecer capacidad de innovación.
- Aspirantes (*challengers*). Aquellos que dominan el momento actual, tanto en tecnología como en despliegue, pero no son parte activa de la tendencia.

Grupos del Cuadrante mágico.



Fuente: [www.gartner.com](http://www.gartner.com).

En el Cuadrante mágico de Gartner presentado en el 2014 sobre analítica de negocio y plataformas analíticas queda patente la relevancia de los datos geográficos para los sistemas de BI. De hecho, cuatro de los cinco criterios incorporados por Gartner (los cuatro primeros) para ponderar la tendencia y actualidad del mercado están estrechamente relacionados con información geográfica:

- Geoespacial e inteligencia de localización.
- Análisis avanzados embebidos.
- *Mashup* y modelado de datos.
- Análisis embebido.
- Soporte de *big data*.

El presente documento cubre las tendencias de las soluciones de localización en entornos de negocio.

Según Gartner, si se extrae la capacidad de los 17 puntos que se usan para evaluar el presente y la tendencia de los sistemas de negocio, se obtiene que, como mínimo, los siguientes puntos tienen vinculación directa con los datos geográficos:

- **Informes.** Proveer de la capacidad de definir un formato de informe para impresión o análisis interactivo.
- **Cuadros de mando (*dashboards*).** Podríamos ver el cuadro de mando como un informe o panel gráfico interactivo para representar las variables de rendimiento. Incluye capacidad de publicar varios objetos e informes interactivos y filtrar mediante parámetros la información mostrada en estos. Los cuadros de mando a menudo emplean componentes visuales, entre ellos mapas.
- **Informes y consultas *ad hoc*.** Permite a los usuarios interrogar los datos sin depender de terceros y crear informes con los datos seleccionados. En particular, las herramientas deben tener una capa semántica reutilizable para permitir a los usuarios navegar por las fuentes de datos disponibles, métricas predefinidas, jerarquías, etc.
- **Visualización interactiva.** Permite la exploración de datos mediante la manipulación de los elementos visualizados (imágenes, gráficos, mapas de calor, mapas geográficos, etc.) que representan los aspectos del conjunto de datos que se están analizando.
- **Localización geoespacial e “inteligente”.** Admite visualizaciones y análisis proporcionados por un contexto geográfico, espacial y temporal. Permite combinar datos geográficos con información de terceros (mapas satélites, datos socioeconómicos) o datos empresariales. Esto le permite representar sus patrones físicos, estableciendo relaciones entre datos e incorporando algoritmos geoespaciales (superposición de datos en mapas interactivos, cálculos de distancia, cálculo de rutas, *geofencing*<sup>2</sup>, etc.) y realizar visualizaciones más usables (utilizando mapas de calor, mapas temporales, agrupaciones de datos o *clustering* o visualizaciones en tres dimensiones, por ejemplo).

<sup>(2)</sup>El *geofencing* es un evento geográfico. Un caso típico de uso es la definición de un área en el mapa que se quiera controlar (por ejemplo, zona reservada), de tal modo que si una entidad entra, sale o cruza el área delimitada (por ejemplo, un coche), se dispara un evento con su consecuente acción (por ejemplo, una notificación).

Las evidencias presentadas apuntan al hecho de que la información geográfica y su manipulación/análisis son relevantes en el mundo del BI y uno de los principales problemas que hoy en día hay que resolver.

## 2. Sistemas de información geográfica

Desde hace siglos la humanidad ha tenido la necesidad de representar mapas en un medio, ya sea sobre piedra o madera, o mediante el papiro, el papel o los actuales formatos digitales.

Uno de los primeros mapas de los que se tiene conocimiento es una tablilla de arcilla de Babilonia de hace unos 5.000 años, que se expone en el Museo Británico de Londres. Se considera que los babilonios utilizaban mapas para medir distancias desde el año 2.300 a. C. con la intención de recaudar impuestos.

Primer mapa babilonio del mundo.



Fuente: <http://www.cubadebate.cu/noticias/2010/02/19/primer-babilonio-mapa-mundo/#.U02dweYXTPA>.

Sin embargo, con el paso de los años los mapas han ido evolucionando técnicamente, en formato, en calidad y en exactitud. Ahora no solo representan la realidad como una fotografía fija, sino que también la modelan en representación y comportamiento. Esto permite interrogar activamente los elementos contenidos en un mapa, siendo posible realizar preguntas como las que se muestran a continuación.

- ¿Qué pasaría si...?
- ¿Existe alguna relación entre...?

- ¿Y si cruzo esta información con...?
- ¿Existe algún modelo similar para...?

Los SIG pueden definirse de muchas maneras, según la óptica desde la que se traten. Una de las definiciones más extendidas es la de Kjerne y Deuker:

Un sistema de hardware, software, datos, personas, organizaciones y convenios institucionales para la recopilación, el almacenamiento, el análisis y la distribución de información de territorios de la Tierra.

No obstante, según la web del Instituto Geográfico Nacional (IGN), la definición más extendida es la que se puede extraer del propio Departamento de Medio Ambiente (DOE) y de Burrough, Goodchild o Rhin:

Conjunto integrado de medios y métodos informáticos, capaz de recoger, verificar, almacenar, gestionar, actualizar, manipular, recuperar, transformar, analizar, mostrar y transferir datos espacialmente referidos a la Tierra.

En la misma web, completan la definición con la de Burrough y Bouillé, en la que se afirma que un SIG debe verse también como un modelo del mundo real. Según esto, un SIG podría definirse como:

Modelo informatizado del mundo real, en un sistema de referencia ligado a la Tierra para satisfacer unas necesidades de información concretas.

En este documento asumiremos como válida la acepción del concepto SIG definida por Burrough, Goodchild o Rhin.

#### Referencia bibliográfica

K. J. Dueker; D. Kjerne (1989). *Multipurpose cadastre: Terms and definitions*. Falls Church, Virginia: American Society for Photography and Remote Sensing and American Congress on Surveying and Mapping.

### 3. Los SIG en los procesos de analítica de negocio

Los sistemas de inteligencia de negocio ofrecen un valioso conocimiento del estado del negocio y una detallada predicción de acontecimientos futuros. Sin embargo, sin el valor del componente geográfico ofrecen una visión parcial, pues no identifican el contexto geográfico de los datos.

Los SIG se orientan a distintos ámbitos y finalidades, como podrían ser la elaboración de mapas en un instituto cartográfico o el análisis de canalizaciones en una red de distribución de aguas.

Este documento se centra en su uso para el análisis del negocio, conocido con los nombres de *geomarketing* y *location analytics*, entre otros. Según Chasco, podemos definir *geomarketing* como:

“un conjunto de técnicas que permiten analizar la realidad económico-social desde un punto de vista geográfico, mediante instrumentos cartográficos y herramientas de la estadística espacial”.

Coro Chasco (2010). *Métodos de Geomarketing*. Madrid: Instituto Lawrence R. Klein.

Mientras que *location analytics* habilita las capacidades espaciales de los SIG (mapas y herramientas de análisis) al servicio de un mejor conocimiento del negocio.

Ambas definiciones se solapan en cuanto a objetivo y funcionalidades, y lo que las diferencia es el punto de vista del término. *Geomarketing* se centra en la incorporación del dato en el SIG, mientras que las soluciones de *location analytics* son no intrusivas, se centran en la incorporación del SIG en los sistemas de negocio (es el SIG el que se adapta al BI y no el BI el que se adapta al SIG). Ambos términos se engloban bajo el concepto de *geographic business intelligence*, un término mucho más genérico de la misma idea.

#### **Location analytics**

*Location analytics* es un concepto de Esri Inc., empresa del sector SIG.

#### 3.1. Público objetivo y beneficios

El público objetivo de las soluciones de negocio georreferenciadas sería cualquier empresa que quiera tomar decisiones considerando variables del territorio. Entre los beneficios de este tipo de sistemas podemos destacar:

- Optimización de la inversión en acciones de mercadotecnia.
- Un mayor conocimiento del mercado y la habilidad de focalizar esfuerzos en determinados segmentos del mercado.



- Diseñar zonas de ventas, rutas de distribución, rutas de mercadotecnia y rutas comerciales.
- Visualizar los datos de negocio en distintos niveles de agregación geográficos (zonas de ventas, sectores, parroquias, municipios, poblaciones, estados y naciones).
- Identificar puntos de venta, oficinas, sucursales, distribuidores, competencia, etc.
- Localizar oficinas más cercanas, análisis de rutas óptimas y alternativas.
- Determinar el área de influencia para precisar la población que se está cubriendo.
- Responder a preguntas como: ¿es óptima la localización actual de un negocio? ¿Dónde se podría ubicar una nueva sucursal? ¿Dónde dirigir una campaña publicitaria?
- Localización de zonas con un alto potencial de mercado.
- Añadir valor en procesos de marketing directo o de atención al cliente.

## 4. Los datos geográficos

La unidad básica de la representación espacial es el **dato geográfico**, que identifica el lugar en el que se sitúa una determinada ocurrencia.

Los datos son la representación concreta de hechos y constituyen el antecedente necesario para el conocimiento de un fenómeno. La información almacenada en una base de datos se obtiene para una finalidad determinada y es fruto de un proceso interpretativo del usuario.

D. Comas; E. Ruiz (1993). *Fundamentos de los Sistemas de Información Geográfica*. Barcelona: Ariel.

Los datos geográficos presentan tres tipos de componentes:

- El **componente espacial** (localización del elemento/evento descrito).
- El **componente temático** (los datos que describen el elemento/evento).
- El **componente temporal** (ocurrencia y validez del elemento/evento descrito).

### Lectura complementaria

R. Olivella; J. Rodríguez (2011). *Introducción a los sistemas de información geográfica. Conceptos y operaciones fundamentales*. Barcelona: Editorial UOC.

A continuación veremos con más detalle estos componentes.

### 4.1. Componente espacial

El componente espacial define la información asociada a la localización. Esta información incluye su situación física (también denominada localización) e información sobre su sistema de representación, sus atributos específicamente espaciales y la relación con los otros datos.

#### 4.1.1. Localización

La localización de un dato sobre un mapa depende del sistema de representación o **proyección cartográfica** que se utilice para proyectar los elementos de una superficie esférica (como es la Tierra) sobre una superficie plana (como suele ser un mapa). Asimismo, debemos tener en cuenta también el **sistema de referencia** que determinará cómo se ubica cualquier elemento sobre un mapa a partir de un **sistema de coordenadas**.

Estos conceptos son suficientemente complejos para determinar una ciencia propia, la geodesia. A continuación introducimos algunos aspectos generales sobre los elementos necesarios para tener una idea básica de qué son y cómo se gestionan los datos espaciales. En particular, veremos qué son las proyecciones cartográficas, para qué sirven y cuáles son las más extendidas. Veremos

también qué son los sistemas de coordenadas y cuáles son los más utilizados (latitud y longitud). Y, finalmente, qué son los sistemas de referencia y cuáles son los más utilizados actualmente.

### Proyección cartográfica

Una proyección cartográfica permite representar la localización situada en el elipsoide de la Tierra (en 3D) en una superficie plana (un mapa 2D). Las proyecciones facilitan una representación visual más cómoda de datos geográficos en medios de dos dimensiones, como pueden ser un monitor o en papel.

Según la superficie sobre la que se proyecta, se puede definir como proyección cónica, cilíndrica y plana o acimutal (o polar si la tangente es en los polos).

La siguiente figura muestra tres tipos de proyecciones: cónica, cilíndrica y polar. Se puede observar que la elección del tipo de proyección ideal depende de la zona que se quiera representar y qué factores (área, distancia o dirección) son críticos. Por ejemplo, la más habitual, la proyección cilíndrica, ofrece una gran distorsión en los polos, por lo que en el caso de la representación polar la mejor opción es la proyección plana o polar.

Infografía que ilustra las opciones de proyección

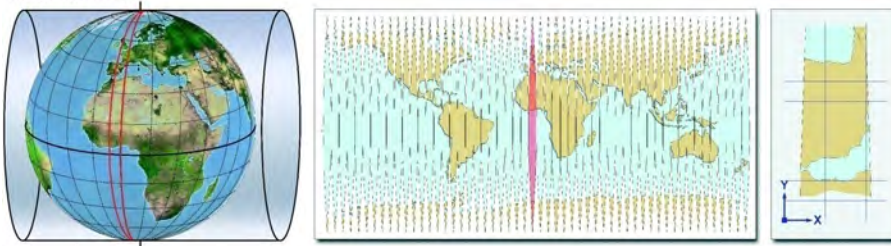


Fuente: <http://www.cartovirtual.es>.

Actualmente, una de las proyecciones más extendidas a nivel mundial es la **Universal Transversal Mercator (UTM)**, o **sistema UTM**, que también se utiliza como sistema de coordenadas. Se trata de un sistema creado por el ejército de Estados Unidos, basado en una proyección cilíndrica, pero que en lugar de realizar la proyección cilíndrica en la tangente del ecuador la realiza en el meridiano.

## Proyección UTM y distribución de husos

### Proyección UTM



Fuente: <http://www.atlasdemurcia.com/>.

La proyección UTM que corta un meridiano permite representar la información cercana al meridiano con un margen de error mínimo. Este margen de error crece a medida que nos alejamos del meridiano de corte. Por este motivo, el sistema de coordenadas UTM divide la Tierra en 60 zonas. A cada zona se la llama huso (por ejemplo, España está incluida en las zonas o husos 28, 29, 30 y 31).

La siguiente figura muestra los distintos husos en los que se divide la península Ibérica. Podemos observar cómo en la parte oeste se utiliza el huso 29, en la parte central el 30 y en Cataluña y Baleares el 31.

Los husos UTM en la península Ibérica



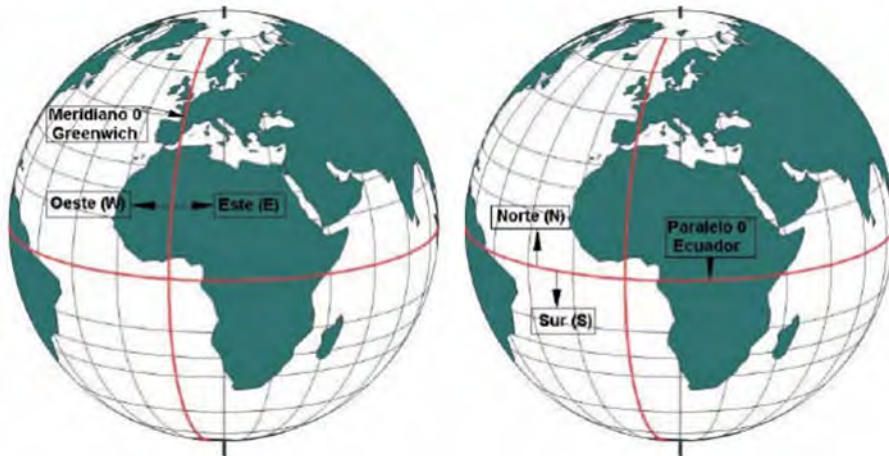
Fuente: <http://ramonortiz1946.wordpress.com/>.

## Sistema de coordenadas

Si bien existen diferentes sistemas de coordenadas, en este caso nos centraremos en el sistema de coordenadas determinado por la proyección UTM y en las coordenadas geográficas. Las coordenadas geográficas permiten ubicar

cualquier elemento en la superficie de la Tierra a partir de su concepción tri-dimensional, y teniendo en cuenta los meridianos (usando el de Greenwich como referencia central) y paralelos (con referencia central en el de Ecuador).

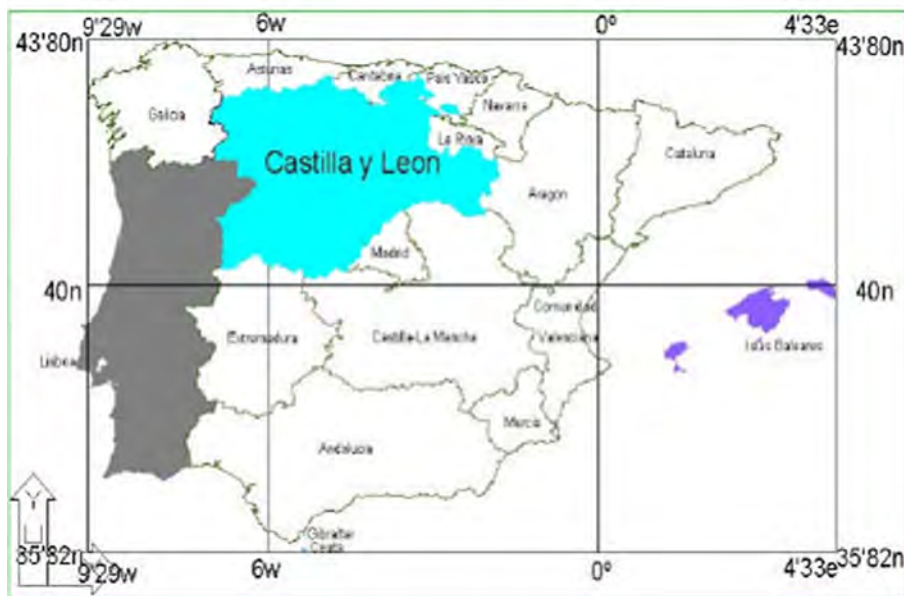
Meridianos y paralelos que permiten determinar las coordenadas geográficas



Fuente: <http://www.cartesia.org/data/apuntes/cartografia/cartografia-geograficas.pdf>.

A partir de la red de paralelos y meridianos y el ángulo que se genera entre ambos, se puede determinar la longitud y latitud de un elemento. De esta manera, la longitud se define por la distancia desde un punto cualquiera de la Tierra al Ecuador, pudiendo tener valores desde 0° hasta los 90° tanto en dirección norte como sur. Por su parte, la latitud se determinaría por la distancia desde un punto cualquiera de la Tierra al meridiano de Greenwich, y los valores que se obtendrían serían de 0° hasta 180° en dirección este u oeste. En la siguiente figura podemos ver las latitudes y longitudes que delimitan la península Ibérica.

Distribución de las coordenadas geográficas en la península Ibérica. Latitud mínima y máxima de la península Ibérica: 35,82°N a 43,80°N. Longitud mínima y máxima de la península Ibérica: 9,29°W a 4,33°E



Fuente: <http://www.cartesia.org/data/apuntes/cartografia/cartografia-geograficas.pdf>.

## Sistemas de referencia

En el apartado anterior hemos visto un par de sistemas de coordenadas (UTM y geográficas). Pues bien, todos los sistemas de coordenadas requieren un sistema de referencia que permita concretar la posición de un elemento en la Tierra con la máxima exactitud teniendo en cuenta la forma irregular del planeta. De este modo, un sistema de referencia vendrá determinado por una serie de constantes físicas en el territorio que nos permitan el cálculo de la posición respecto al sistema de coordenadas elegido; esto es lo que se denomina *datum*. Actualmente podemos determinar dos sistemas de referencia como los más utilizados:

- **WGC84.** Se trata de un sistema de referencia a nivel internacional y utilizado por ejemplo por los dispositivos GPS.
- **ETRS89.** Se trata del sistema de referencia oficial en Europa y permite una mayor precisión en el territorio europeo que el WGC84.

### 4.1.2. Modelos de representación

Una vez que sabemos cómo describir datos geográficos, el siguiente paso es ver en qué formato se pueden almacenar. Se pueden distinguir principalmente dos tipos de representación: *vectorial* y *raster*. El uso de una u otra dependerá de si los valores de los datos que hay que representar son discretos (representación *vectorial*) o continuos (representación *raster*).

En la siguiente figura podemos ver la misma realidad con dos representaciones distintas; en la parte izquierda el modelo *vectorial* junto a su tabla de almacenamiento (con tres valores que se corresponden a las tres zonas representadas) y en la parte derecha la representación *raster*, donde cada celda almacena su propio valor. Como consecuencia, el espacio requerido para almacenar datos *raster* tiende a ser significativamente mayor.



Vectorial y raster, dos modos de representar la misma realidad

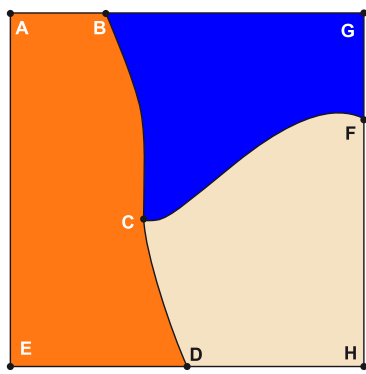


Imagen vectorial

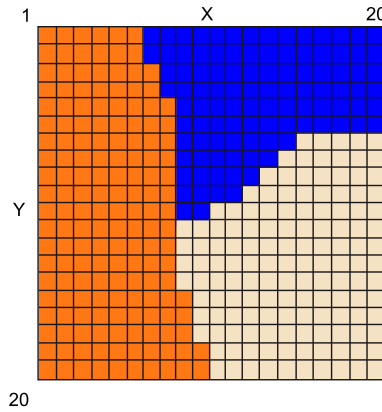


Imagen de trama

Vector		
ID de polígono	Coordenadas	Tipo de tierra
1	A, B, C, D, E	Caliza
2	B, C, F, G	Arcilla
3	C, F, H, D	Grava

Trama	
Ref. cuadrícula	Elemento
X=1, y=1	Caliza
X=2, y=1	Caliza
X=3, y=1	Caliza
X=4, ... etc.	...
X=20, y=20	Grava

Fuente: <http://civilgeeks.com/>.

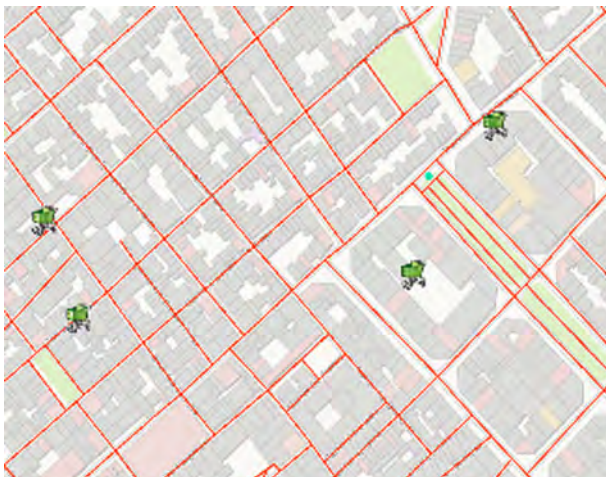
### Modelo vectorial

En el modelo vectorial de los datos, los elementos básicos son puntos y segmentos rectos. Un punto se representa mediante un par de coordenadas (X e Y). Un segmento es una línea recta que une dos puntos.

Conjugando este par de elementos, se pueden representar los siguientes elementos:

- **Puntos:** elementos localizados por una X e Y.
- **Líneas:** segmentos definidos a partir de las coordenadas de puntos.
- **Polígonos:** superficies lineales cerradas.

Almacenes, ejes de las tramas de la red de calles y edificios de Barcelona



Fuente: Elaboración propia. Mapa base Esri.

En la imagen anterior se pueden observar los elementos del modelo vectorial:

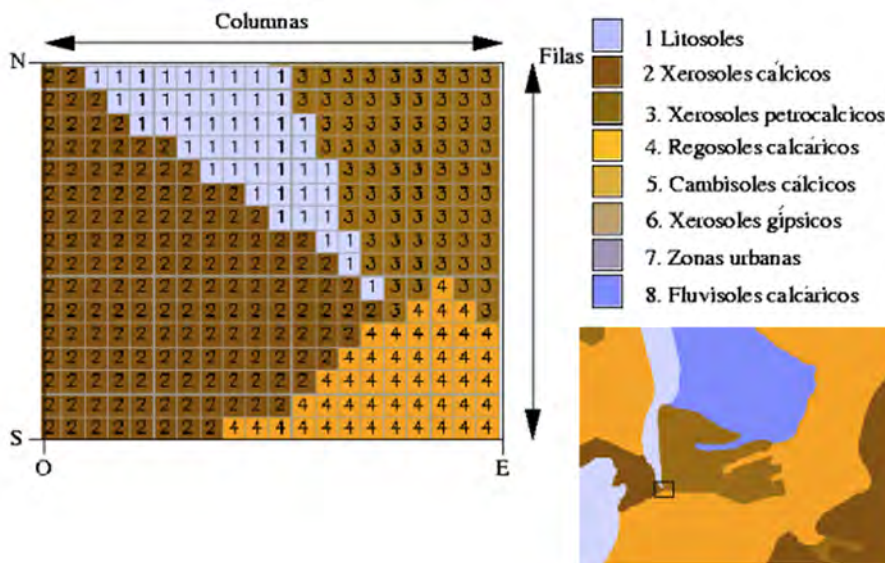
- **Puntos:** almacenes representados por el carrito de la compra.
- **Líneas:** representados por las líneas rojas que muestran el eje de la red de calles.
- **Polígonos:** edificios de color crema.

**Modelo raster**

Una imagen tipo *raster* está definida por una malla de celdas donde cada una de ellas representa un valor. La definición de cómo se define la celda y sus posibles valores dependerán de lo que se quiera representar.

La siguiente figura muestra un modelo de representación *raster* de usos del suelo. En la izquierda hay un detalle de la sección de la malla de localización con la información asociada a cada valor en la celda; dicho valor establece la categoría del uso del suelo.

Ejemplo de representación cualitativa del modelo raster



Fuente: SIGMUR (Universidad de Murcia).

El ejemplo podría ser útil en el caso de una empresa de venta de aparatos de descalcificación de agua. Se podrían identificar zonas urbanas cerca de suelos cálcicos, candidatas a una determinada acción de marketing comercial.



### 4.1.3. Atributos espaciales

Para que los elementos espaciales definan la realidad de manera fiable, deben proveer ciertos atributos referentes a su localización que permitan conocer el rigor de los datos y otros aspectos adicionales, como por ejemplo el área de un edificio o la longitud de una carretera.

Estos atributos están sujetos al modelo utilizado para la representación y explicado anteriormente en este mismo documento. Algunas de estas propiedades para los modelos vectoriales y *raster* son:

- Modelo vectorial:
  - Puntos: sistema de coordenadas.
  - Líneas: sistema de coordenadas, medida de longitud, forma, pendiente y orientación.
  - Polígonos: sistema de coordenadas, área, perímetro, forma, pendiente y orientación.
- Modelo *raster*:
  - Malla de celdas: la unidad mínima o celdas son representaciones poligonales, por lo que sus propiedades son: sistema de coordenadas, área, perímetro, forma, pendiente y orientación.

### 4.1.4. Relaciones espaciales

En un SIG los datos representan la realidad y esta realidad se estructura a modo de capas de información. Por ejemplo, un modelo de usos del suelo sería la capa de “usos del suelo”, mientras que un modelo que muestre un tendido eléctrico podría ser una capa “cable de red”; la visualización de las torres de soporte de la línea eléctrica podría integrarse en una capa llamada “torres”.

Un SIG modela la realidad y en la realidad los elementos se relacionan y se ordenan por reglas, ya sean **físicas**, una “torre” del tendido eléctrico soporta “cable de red”, por lo que la capa “cable de red” debe estar por encima de la capa “torres”; ya sean **legales**: un tendido eléctrico de alto voltaje (capa “cable de red”), por ejemplo, no puede pasar por encima de un área protegida (“capa usos del suelo”).

Estas relaciones o reglas entre los elementos pueden definirse mediante un campo común entre las dos capas. La capa “torres” y la capa “cable de red” pueden tener un identificador de línea eléctrica común, al igual que se establece en una base de datos relacional (conocida habitualmente como **relación por campo atributo o join**).

Sin embargo, no siempre dos capas tienen un campo común; por ejemplo, en el caso de la relación entre el tendido eléctrico y un área protegida. En este caso, el tipo de relación entre las dos capas es espacial: el tendido eléctrico no puede tocar un área protegida, en lo que se conoce como **relación espacial** o *spatial join*.

La capacidad de poder establecer **relaciones espaciales** es una de las características más importantes que justifica un SIG. Sin la información geográfica almacenada, es imposible poder establecer este tipo de relaciones.

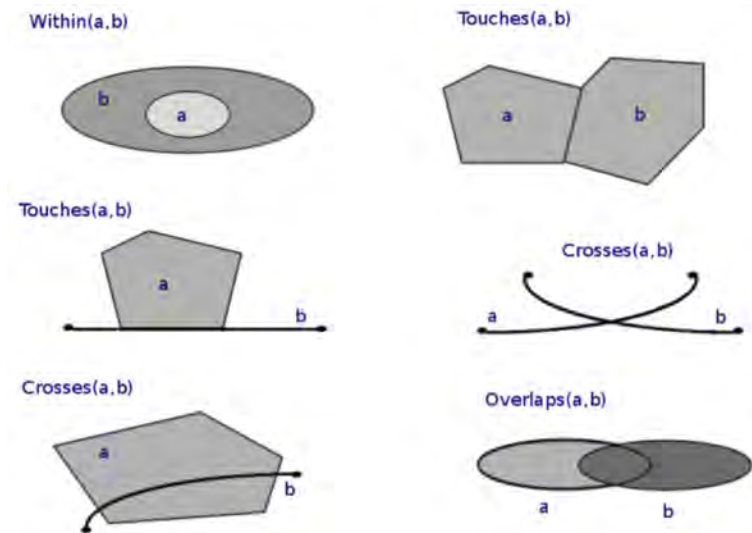
Una **relación espacial** nos permite relacionar información tan dispar (y lógica) como: “una capa de locales disponibles cerca de una carretera principal, a menos de 5 minutos en coche de un núcleo urbano que cuente con una población superior a las 25.000 unidades familiares, en las que los padres tengan entre 25 y 45 años y un poder adquisitivo medio-alto”. Estas características nos permitirían encontrar una buena ubicación para un concesionario de coches especializado en la venta de utilitarios familiares.

Una relación permite comprobar distintas propiedades geográficas, por ejemplo:

- Si un elemento está contenido en otro.
- Si un elemento coincide con otro.
- Si un elemento se interseca con otro.
- Si un elemento se superpone a otro.
- Si un elemento toca a otro.
- Si un elemento cruza a otro.
- La distancia de uno respecto al otro.
- Etc.

En el siguiente ejemplo, se pueden observar varios tipos de relaciones entre varios elementos de distinto tipo: puntos, líneas y polígonos.

## Ejemplo de relaciones espaciales



Fuente: Wikipedia.

Una relación espacial no siempre se utiliza en consultas; existen casos en los que las relaciones entre los elementos establecen las restricciones de integridad del modelo de datos.

Las **relaciones topológicas** son un conjunto de reglas que definen cómo coexisten en el mismo espacio distintas capas. Se trata de relaciones cualitativas. Un ejemplo de ello puede ser una zona construida que deba formar parte de una zona urbana. O, como en el ejemplo anterior, una torre eléctrica situada dentro de una línea de tendido eléctrico.

Las relaciones topológicas establecen cómo deben evolucionar los elementos geográficos al modificar alguno de sus elementos relacionados. Por tanto, al modificar uno de los elementos de las capas, se deben utilizar las relaciones topológicas para rectificar los elementos en los que están relacionados.

Como podemos ver en la siguiente figura, los límites administrativos de los países son un claro ejemplo de reglas topológicas. Si un país redefine una frontera, automáticamente quedan modificados los límites de los países colindantes.

Mapa de político mundial



Fuente: <http://www.osgeo.org/>.

## 4.2. Componente temático

No es lo mismo visualizar una capa con los nombres de los barrios que una capa con las ventas anuales para cada uno de ellos. Todo elemento geográfico tiene información alfanumérica asociada que puede ser utilizada en función de qué punto de vista de la realidad se quiere mostrar.

El componente temático, también conocido como atributos o variables, son los valores o campos (cuantitativos o cualitativos) que definen una capa y que definen la naturaleza y el objetivo de la información mostrada.

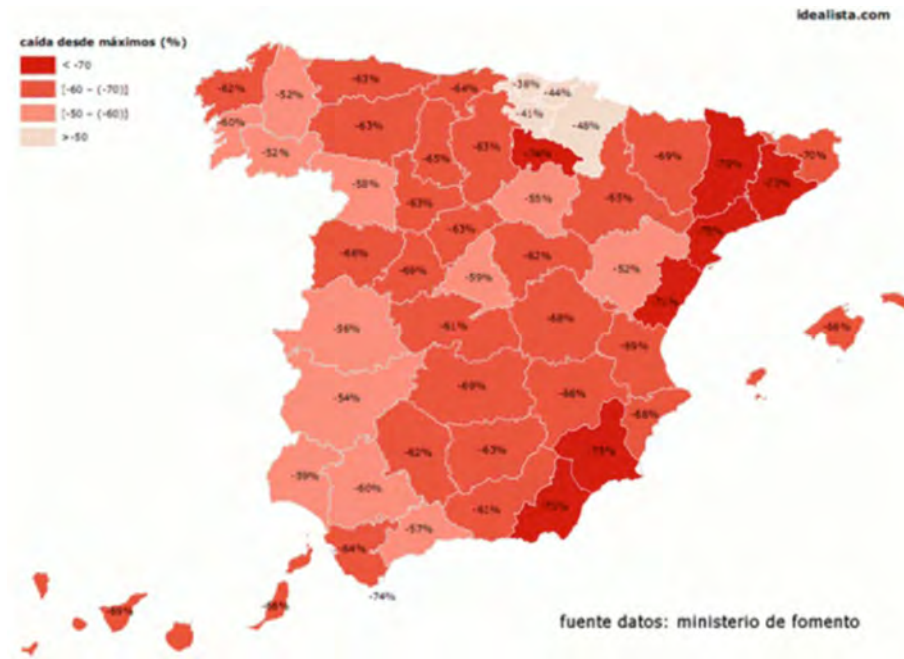
Este componente temático depende del modelo de datos utilizado y de la naturaleza del valor.

### 4.2.1. Clasificación en función del modelo

En el **modelo vectorial**, el componente temático de los elementos viene dado por los datos asociados a la geometría. Estos elementos deben ser recogidos explícitamente en función del modelo que se ha de representar. Los atributos suelen ser múltiples, como en el caso de representar provincias, cuando los atributos asociados podrían ser la población, la renta per cápita o los idiomas oficiales.

La siguiente figura muestra un mapa temático que ilustra la caída de precios de las viviendas en el 2011 por provincias en España. Los valores con un rojo más vivo muestran las zonas con una caída más pronunciada (concentradas mayoritariamente en la costa este), mientras que los colores más claros ilustran la zona con una menor caída (en la zona del País Vasco y Navarra).

Mapa temático de la caída de las ventas de vivienda en el 2011 desde máximos por provincias

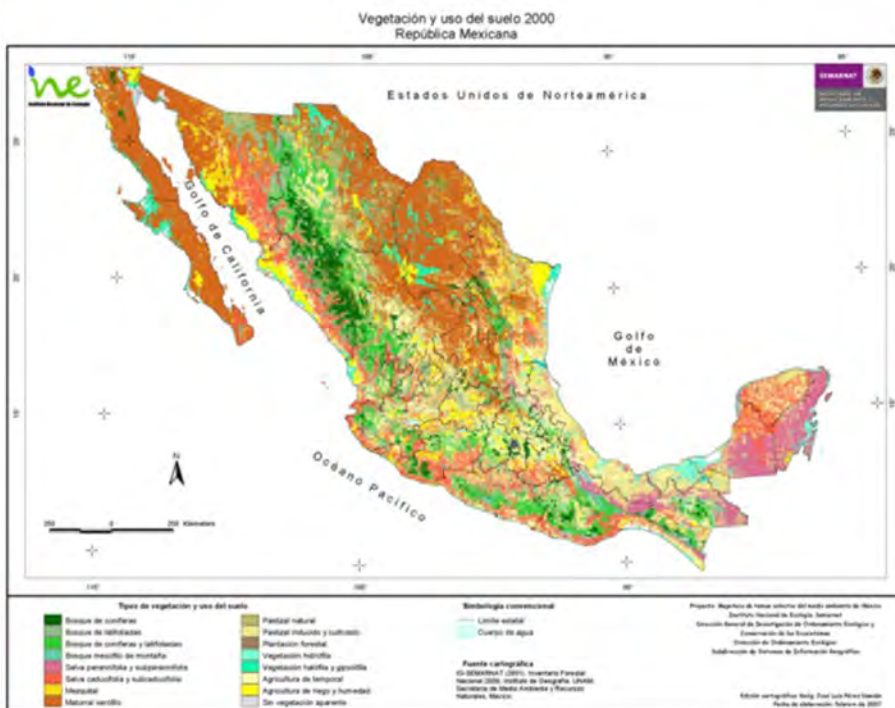


Fuente: Ministerio de Fomento.

En el **modelo raster**, el componente temático está implícitamente ligado dentro de la capa a partir de la celda representada en la malla. El componente temático forma parte del propio valor del *raster* y normalmente se genera al clasificar los valores de la imagen.

La siguiente imagen muestra un mapa temático de usos del suelo basado en un modelo *raster* de México.

Mapa de usos del suelo y vegetación de México (2007)



Fuente: Instituto Nacional de Ecología de México.

#### 4.2.2. Clasificación en función de la naturaleza

Según la naturaleza de los datos que hay que analizar y del tipo de información que contienen, los podemos clasificar en:

- **Cualitativos o cuantitativos:**
  - Los valores **cualitativos** son descriptivos. Un ejemplo de valor cualitativo sería la clasificación de bares en función de su tipo (cafetería, cervecería, bar de copas, etc.). En esta clasificación, las operaciones aritméticas, siempre que no sean de agrupación o igualdad, pierden su sentido.
  - Los valores **cuantitativos** son numéricos y representan una medición, como población, temperatura o valor.
- **Discretos o continuos:**
  - Los valores **discretos** son finitos (un entero o un conjunto de valores, por ejemplo). La siguiente figura ilustra el número de librerías por comunidad autónoma.

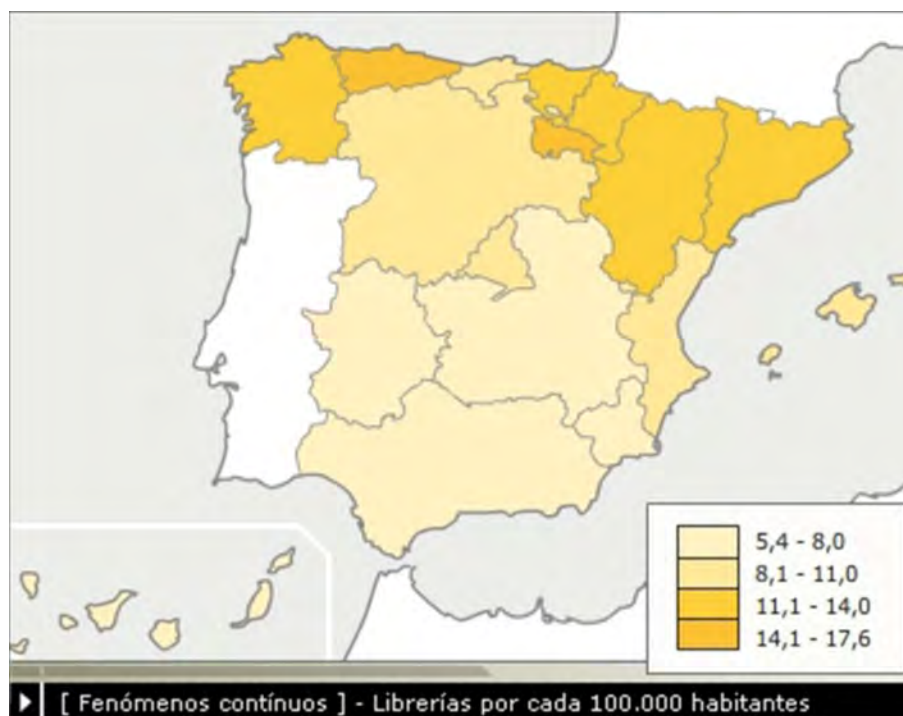
Número de librerías por comunidad autónoma (23/07/2007)



Fuente: Instituto Geográfico Nacional.

- Los valores **continuos** representan un rango continuo infinito de valores. Algunos ejemplos de valores continuos serían la densidad de población de un municipio o las poblaciones colindantes con una gran ciudad, que suelen disminuir su densidad de población a medida que nos alejamos del centro principal.

Densidad de librerías por cada 10.000 habitantes por comunidad autónoma  
(23/07/2007)



Fuente: Instituto Geográfico Nacional.

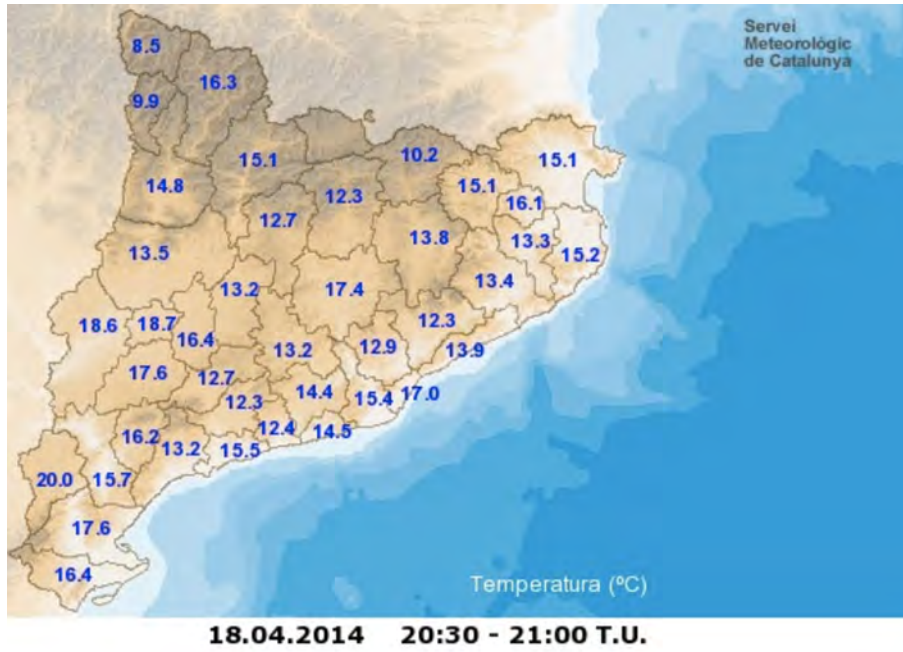
### Valores de correlación

Esta clasificación se realiza en función de la similitud que representan los datos:

- **Autocorrelación espacial.** Son aquellas donde los valores se asemejan entre elementos próximos entre sí. La siguiente imagen muestra los valores de temperaturas en Cataluña el 18 de abril del 2014 entre las 20.30 y 21 horas. Se puede observar cómo no existen saltos bruscos de temperaturas en comarcas con similar orografía, por lo que siguen una correlación espacial similar en el mismo instante.



Mapa de evolución de las temperaturas de Cataluña



Fuente: <http://www.meteo.cat>.

- **Autocorrelación temporal.** Son aquellas en las que los valores se asemejan en función de la proximidad **temporal** en que fueron tomados. En la imagen inferior, se puede observar la evolución del ciclón tropical Gustav y cómo su crecimiento es similar en función de la progresión temporal.  
Evolución de un ciclón tropical Gustav



Fuente: <http://www.meteoraleja.es/>.



### 4.3. Componente temporal

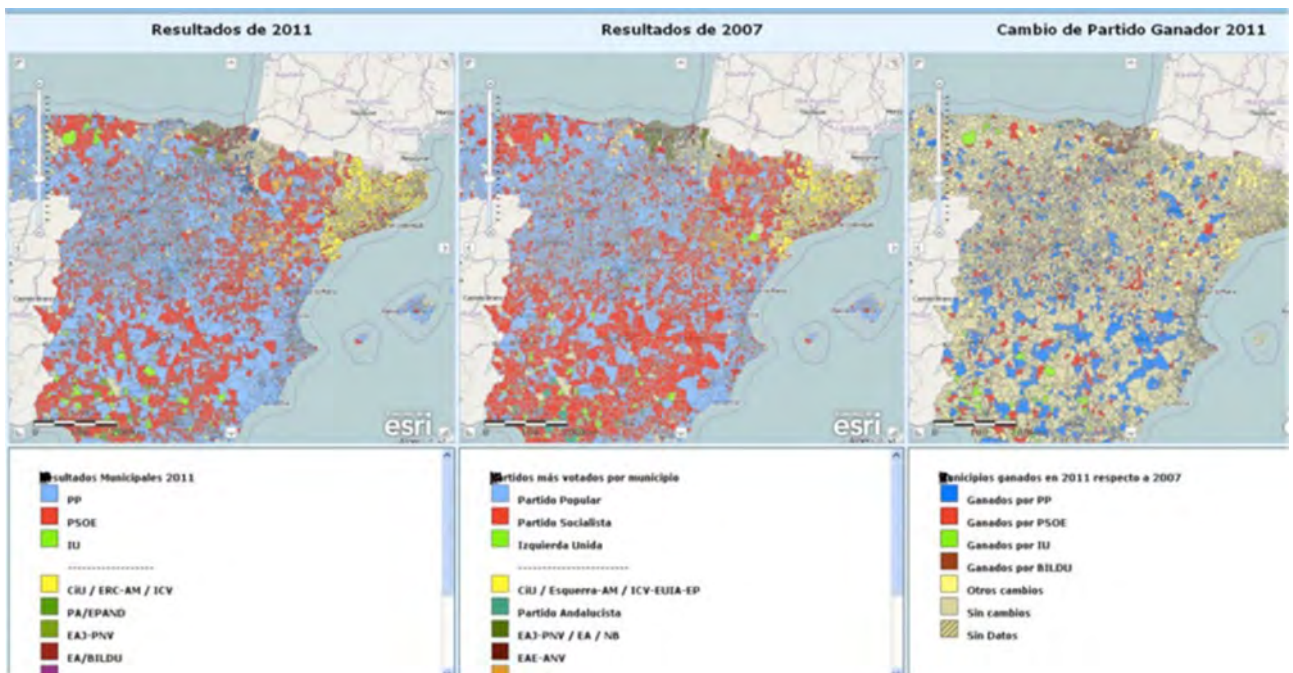
El componente temporal se centra en la representación de la variabilidad de los fenómenos geográficos en el tiempo.

Este componente está íntimamente ligado al componente temático y sirve sobre todo para identificar variables de correlación temporal tanto en posición, por ejemplo la trazabilidad de una flota de vehículos, como en sus atributos, por ejemplo el índice de crecimiento poblacional en un municipio (componente temático).

El componente temporal se suele representar mediante:

- **Mapas comparativos temporales o secuencias de mapas.** Se comparan dos mapas que comparten el mismo espacio geográfico pero distinto espacio temporal.

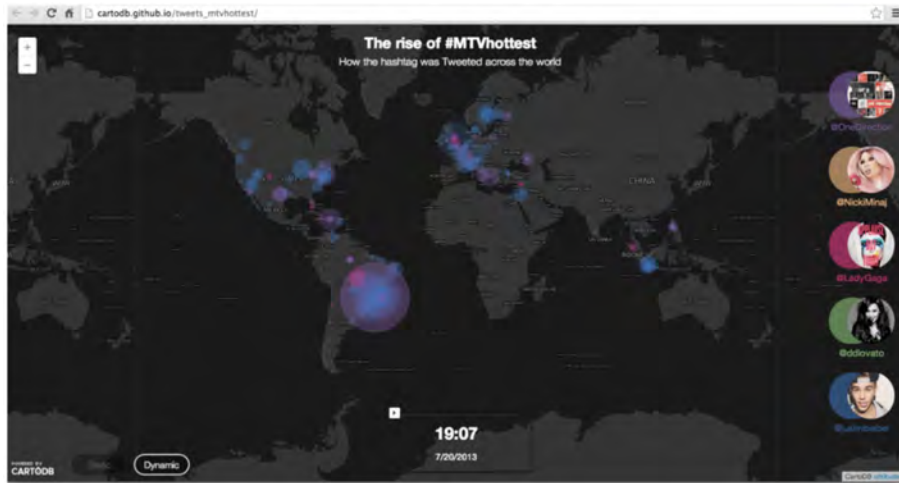
Mapa comparativo de evolución del voto entre los años 2011 y 2007



Fuente: <http://www.votosycifras.com>.

- **Mapa animado.** En un único visor se centra la evolución animada de un determinado fenómeno.

### Seguimiento de tuits relativos a artistas famosos por día y hora



Fuente: <http://cartodb.com/>.

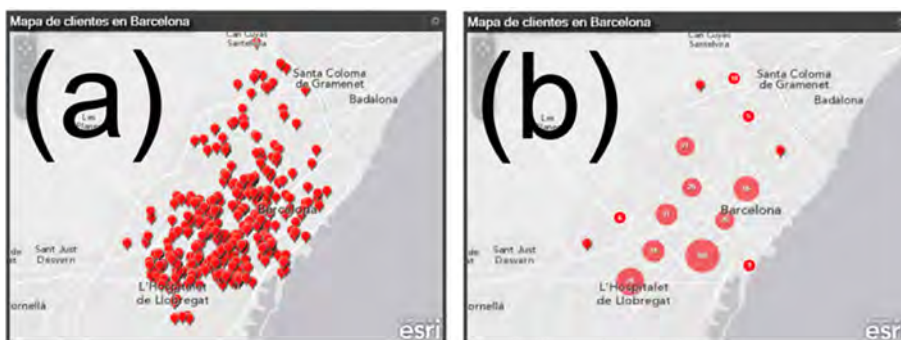
## 5. Comunicación cartográfica

Una vez vistos los componentes relativos a los datos geográficos, en esta sección vamos a estudiar métodos para comunicar la información en un mapa. Los dos siguientes apartados tratan sobre los métodos para ordenar, en caso de ser necesario, los datos en un mapa y acerca de los distintos métodos de representación para que la comunicación sea efectiva.

Uno de los objetivos de un sistema de información geográfica es la correcta visualización de los datos. No basta con representar los datos, sino que estos deben ser representados de una manera coherente y usable que nos permita extraer información e identificar patrones a partir de su visualización.

La siguiente figura pretende representar la distribución de clientes en la zona de Barcelona. Sin embargo, en la representación de la figura (a) es difícil saber dónde está la mayor concentración de clientes, pues la representación está masificada. En la figura (b) se opta por otro tipo de representación para mostrar los mismos datos. Como vemos, usar una representación agrupada sobre estos datos facilita visualizar dónde se halla la mayor concentración de clientes.

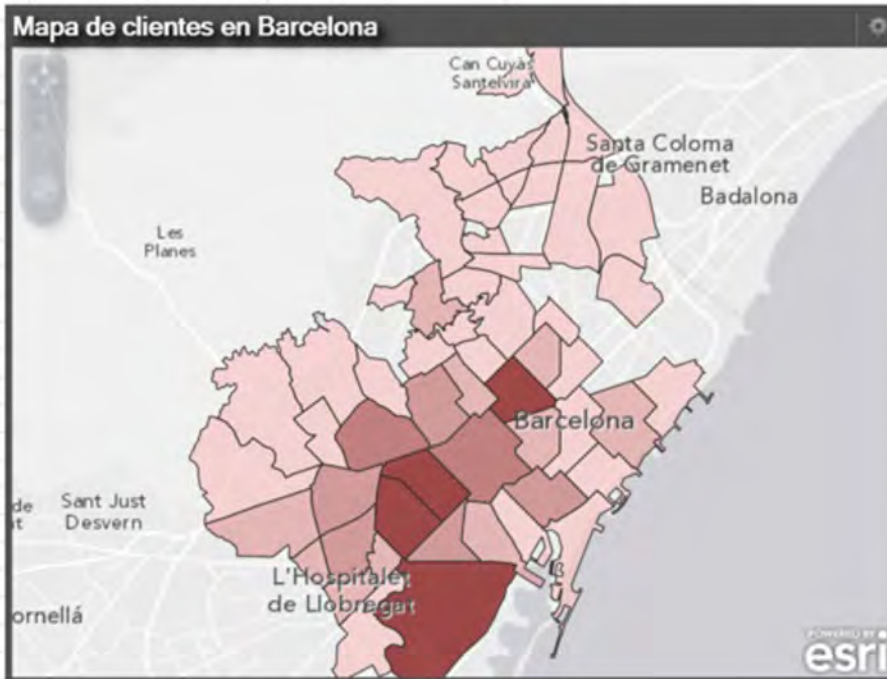
Mapa de puntos de ventas de Barcelona. (a) Representación de elementos puntuales puros. (b) Representación de elementos agrupados



Fuente: Elaboración propia. Mapa base Esri.

Otra alternativa de representación interesante en este caso sería la tematizada por barrios en Barcelona, como puede verse en la figura inferior.

Mapa de puntos de ventas por barrios en Barcelona



Fuente: Elaboración propia. Mapa base Esri.

## 5.1. Métodos de representación cartográfica

Mayoritariamente los métodos de representación cartográfica son comunes en los diferentes tipos de geometría que nos podemos encontrar en un mapa (básicamente, puntos, líneas y polígonos).

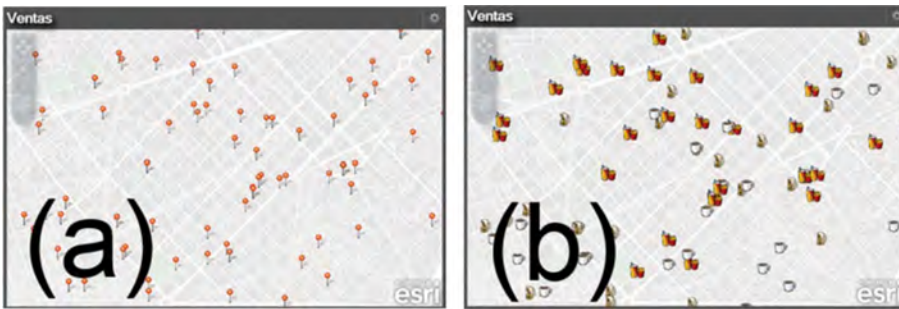
Este apartado describe los principales métodos adoptados en la representación cartográfica, que junto con las técnicas descritas en el apartado posterior cubren la mayoría de las necesidades en el ámbito de la representación de mapas.

### 5.1.1. Representación simple

A menudo no es suficiente mostrar los datos como una representación puntual simple en un mapa. Si la capa que se ha de representar es única, un símbolo geométrico puede ser suficiente. No obstante, cuando los elementos que se deben representar son cuantiosos, es recomendable ver si se pueden tipificar los datos y utilizar representaciones gráficas (iconos) distintas para cada tipo de elemento.

Por ejemplo, ¿cuál de estos mapas muestra la información de una manera más clara? Obviamente, el segundo mapa ofrece más información visualmente, lo que facilita a quien lo analice un mejor entendimiento de los datos que representa.

En el mapa (a) todos los puntos están representados con el mismo icono; en el mapa (b) se distingue por cafetería, cervecería y bar



Fuente: Elaboración propia. Mapa base Esri.

### 5.1.2. Clasificación por categorías

Una categoría clasifica elementos de un mismo grupo. Esta clasificación puede ser por nivel de importancia o valor o por tipo (según naturaleza de la entidad). Por ejemplo, si se analiza la población, podemos establecer categorías según rango de edad o sexo. De esta manera, elementos que tienen valores parecidos formarán parte de la misma categoría.

En los SIG, una categoría define una agrupación de los elementos en una capa. Su utilización en un mapa es muy útil para mostrar la información de una manera efectiva y visual. En el ejemplo anterior, en el mapa (b), los elementos forman parte de la misma capa, llamada "local", y se representan en función de la categoría "tipo de local" (cafetería, cervecería y bar).

Si en la clasificación se utiliza la naturaleza de su valor, una categoría puede venir definida por:

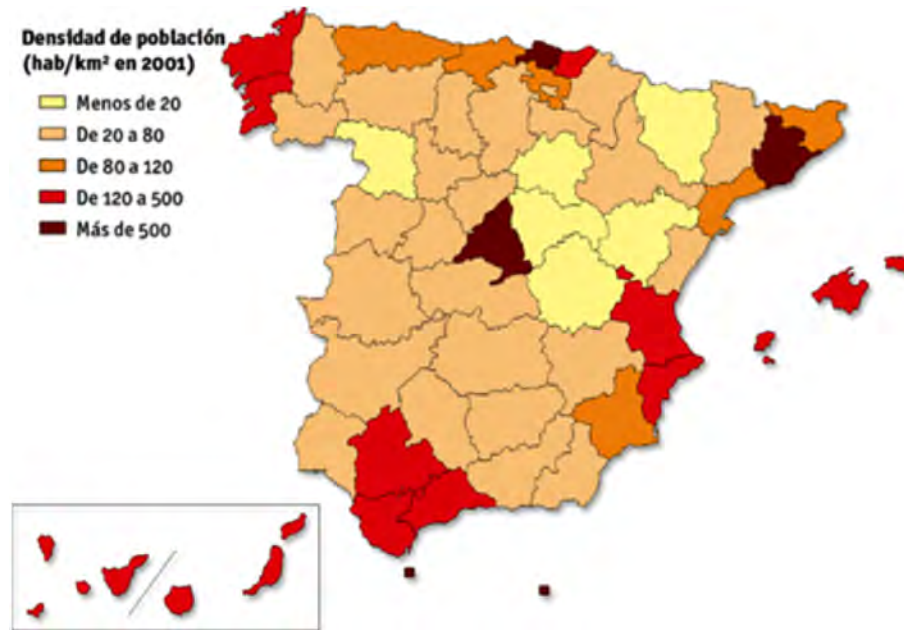
- **Valores cualitativos**, donde el valor representa una cualidad del elemento y describe una tipología, condición o estado. Por ejemplo, el tipo de terreno: zona urbana, zona interurbana, terreno rústico, etc.
- **Valores cuantitativos**, donde la descripción forma parte del propio valor numérico; por ejemplo: la facturación anual.

Si se considera la clasificación en cuanto a agrupación de elementos, una categoría puede definirse por:

- **Valores únicos**, como por ejemplo la clave principal de un elemento.
- **Valores únicos para cada categoría**, como por ejemplo los usos del suelo: forestal, de cultivo o urbano.
- **Clasificación por rangos**, como por ejemplo en el siguiente mapa, donde se han establecido cinco clases para representar la densidad de población:



La densidad de población de España por provincia, según el censo del 2001 del Instituto Nacional de Estadística



Fuente: <http://www.ine.es/>.

En el caso de la clasificación de los valores cualitativos, la elección de estos se deja al criterio del diseñador del mapa. Por ejemplo, si lo que se quiere es un valor único para cada tipo de suelo, una opción podría ser: bosques, matorral, forestal, cultivos libres, cultivos de secano, regadío, zonas urbanas sin urbanizar o zonas urbanas construidas; pero en el caso de querer generalizar los elementos representados se podrían utilizar agrupaciones de atributos, como se muestra en la siguiente figura: terreno forestal (incluye bosques, matorral, etc.), terreno de cultivo (incluye secano o regadío) y terreno urbano (incluye zona sin urbanizar y urbanizada).

Superficie y usos del suelo por comarcas en Cataluña

	Superficie forestal			Conreus			Urbanitzat i altres (1)
	boscos	bosquines	altres	Sense vegetació	secà	regadiu	
Alt Camp	14.393	12.209	:	1.256	18.540	4.060	3.311
Alt Empordà	55.174	24.924	2.284	4.671	25.362	14.966	8.321
Alt Penedès	13.692	13.089	1.761	2.281	22.158	412	5.926
Alt Urgell	91.883	36.841	5.783	2.516	3.507	2.097	2.073
Alta Ribagorça	16.320	22.930	1.518	839	591	180	413
Anoia	29.483	15.271	4.496	2.244	28.891	418	5.947
Bages	65.988	25.131	4.100	4.250	21.958	804	7.729

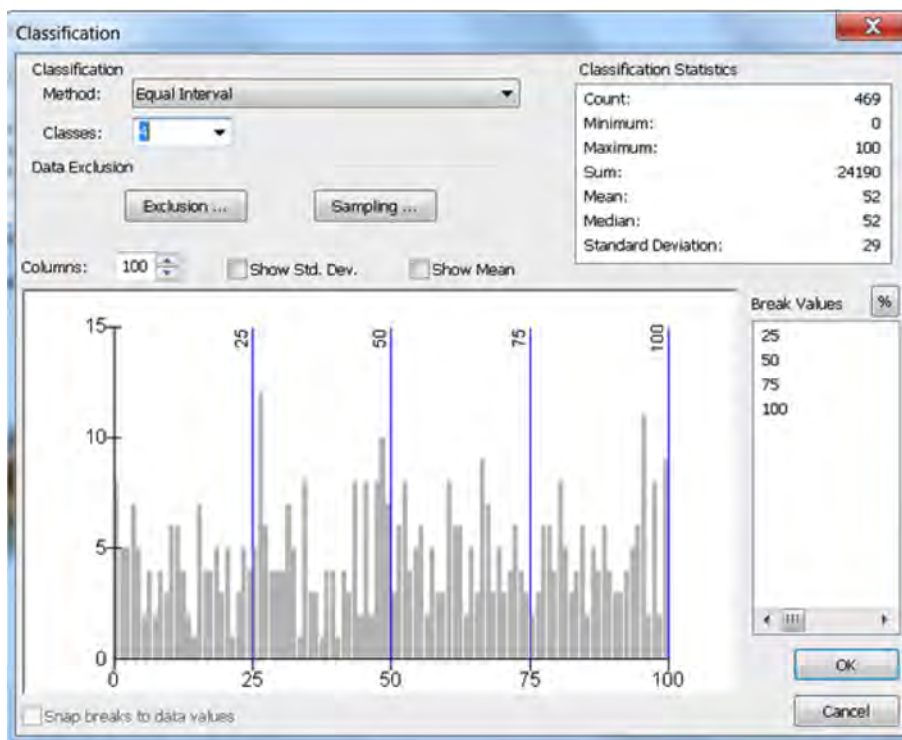
Fuente: <http://www.idescat.cat/>.

En el caso de la clasificación de **valores cualitativos**, el diseñador del mapa también podría establecer los criterios de selección, o por el contrario se podría utilizar alguna de las técnicas de clasificación existente:

1) **Clasificación por intervalos equivalentes**. Se define el número de clases que se quiere obtener y se divide por el valor mínimo y máximo; de esta manera los cortes resultantes son intervalos regulares. Se suele utilizar para la representación de porcentajes o temperatura. Por ejemplo, en el caso de la clasificación del porcentaje de las ventas de clientes en zonas comerciales, si se establecen cuatro clases los valores serán:

- Menor del 25%
- Igual o mayor al 25% y menor del 50%
- Igual o mayor al 50% y menor del 75%
- Igual o mayor al 75%

Clasificación por intervalos equivalentes de 469 elementos con valores entre 0 y 100



Fuente: Captura de ArcMap; elaboración propia.

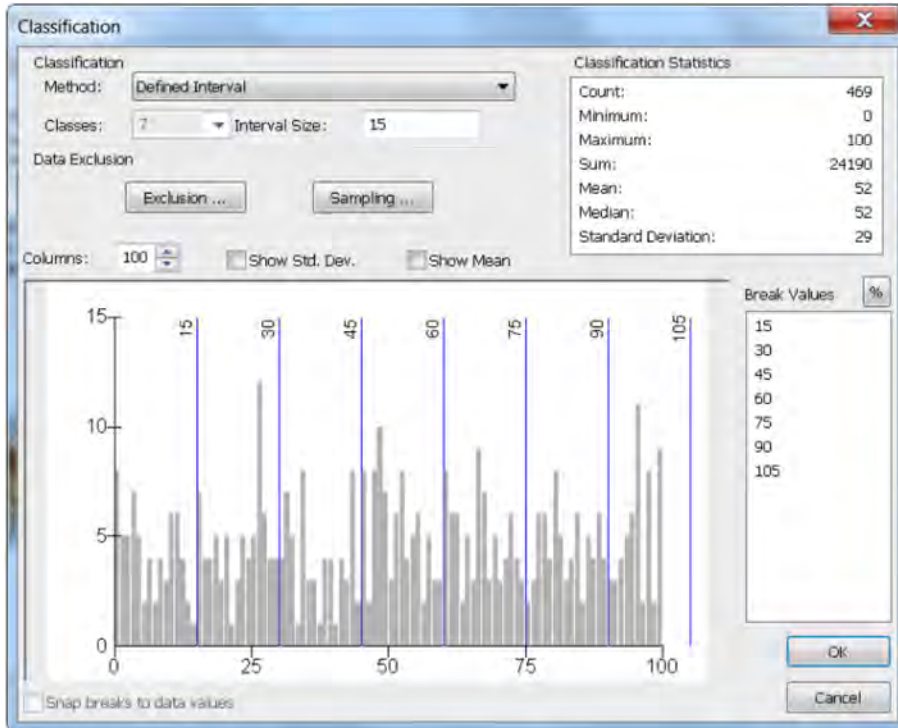
2) **Clasificación por intervalos definidos**. El punto de partida de este tipo de clasificación es la contraria a la técnica usada en el caso de los intervalos equivalentes. En esta clasificación se empieza definiendo las unidades que abarcará cada corte. Por ejemplo, y siguiendo con el caso anterior, si se quiere definir un intervalo de corte del 15%, los cortes resultantes serán:

- Menor del 15%
- Igual o mayor al 15% y menor del 30%
- Igual o mayor al 30% y menor del 45%
- Igual o mayor al 45% y menor del 60%

- Igual o mayor al 60% y menor del 75%
- Igual o mayor al 75% y menor del 90%
- Igual o mayor al 90%

Por lo que se obtiene un total de siete cortes.

Clasificación por intervalos definidos de 469 elementos con valores entre 0 y 100



Fuente: Captura de ArcMap; elaboración propia.

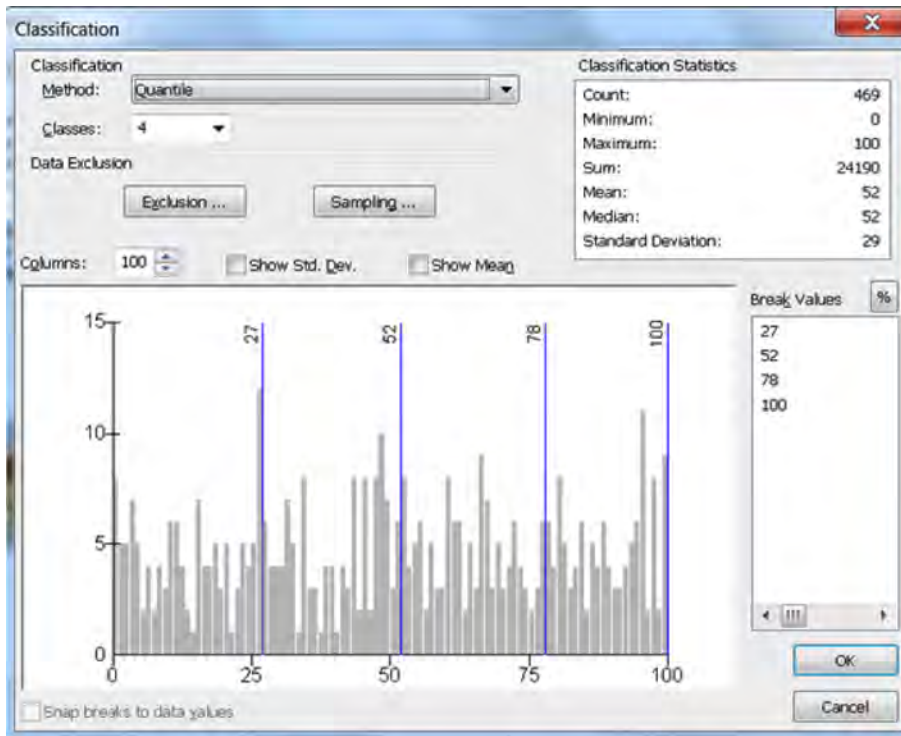
**3) Clasificación por cuantiles.** Se define cuántos cortes se quiere representar y se realiza cada clase con el mismo número de elementos. En esta clasificación no existen los cortes con muchos valores o pocos. Se trata de un método adecuado para la representación de valores lineales.

#### **Nota**

Mientras que la clasificación por intervalos equivalentes realiza los cortes en función de los valores, la clasificación por cuantiles la realiza por número de elementos.



Clasificación por cuantiles de 469 elementos con valores entre 0 y 100

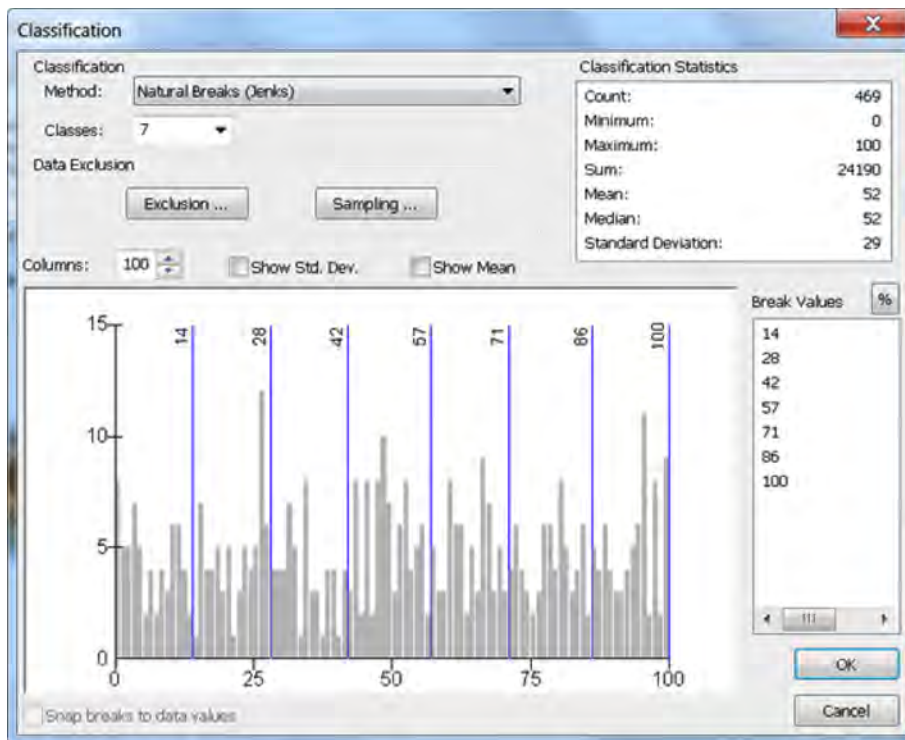


Fuente: Captura de ArcMap; elaboración propia.

**4) Clasificación por cortes naturales o clasificación de Jenks.** Este método no ofrece una distribución lineal de los datos. En esta clasificación se establece el número de cortes que se quiere representar y una función decide cuál es la mejor agrupación para conseguir que cada clase tenga valores similares y, al mismo tiempo, un alto contraste entre ellas.

En la siguiente imagen se observan los valores de rotura utilizando el método de Jenks. Como se observa en el histograma, ofrece una distribución regular en cuanto a los valores entre roturas y el número de puntos contenidos en cada corte.

Clasificación por clasificación de Jenks de 469 elementos con valores entre 0 y 100



Fuente: Captura de ArcMap; elaboración propia.

### Página web

El algoritmo no es trivial: en los siguientes enlaces se pueden encontrar el algoritmo matemático (Object Vision wiki, [http://wiki.objectvision.nl/index.php/Fisher's\\_Natural\\_Breaks\\_Classification](http://wiki.objectvision.nl/index.php/Fisher's_Natural_Breaks_Classification)) y la función desarrollada en python por Daniel Lewis (Volunteered Geographic Information, [http://en.wikipedia.org/wiki/Jenks\\_natural\\_breaks\\_optimization](http://en.wikipedia.org/wiki/Jenks_natural_breaks_optimization)).

## 5.2. Técnicas de representación en mapas

A continuación se definen diferentes técnicas básicas utilizadas para visualizar información en mapas con el fin de incrementar la cantidad y calidad de información que proporcionan al lector.

En este documento solo trataremos algunos de los métodos más comunes y más relevantes en nuestro caso.

### 5.2.1. Agrupaciones o *clustering*

La representación en *cluster* es un sistema de visualización agrupada que clasifica grupos homogéneos de datos que están localizados a distancias relativamente cortas y los agrupa en la vista del mapa.

Las agrupaciones por *clustering*, normalmente, se aplican a elementos con geometrías de tipo punto.

Este tipo de visualización se utiliza en análisis de densidad de puntos (análisis de concentración de puntos por área) para el reconocimiento de patrones en estudios de segmentación de mercado o localización de clientes.

En el ejemplo inferior se puede visualizar dónde está localizada la mayor concentración de clientes en el área de Barcelona.

Mapa de puntos de ventas de Barcelona



Fuente: Elaboración propia. Mapa base Esri.

### 5.2.2. Mapa de calor (*heatmap*)

Un *heatmap* o mapa de calor es un método de representación que se utiliza para identificar zonas significativas en un mapa. Se selecciona un criterio de interés, por posición o por valor, y se representa su densidad sobre el mapa. Por ejemplo, si utilizamos exclusivamente la posición de los clientes, la zona más significativa aparecerá donde exista una mayor concentración de ellos, pero si utilizamos su volumen de gasto, la zona más significativa no tiene por qué ser donde exista la mayor concentración de clientes, sino que será en las zonas en las que realicen el gasto más significativo (aunque sea un punto asilado).


Al final, las zonas con una concentración mayor se corresponden con las zonas con un mayor valor del criterio seleccionado.

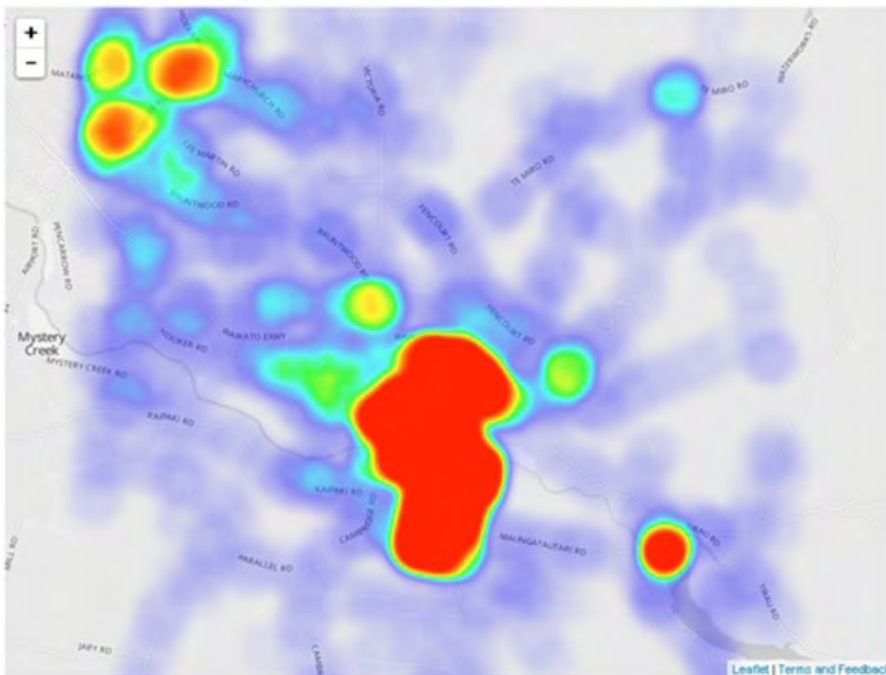
Normalmente, en este tipo de mapas se utilizan colores cálidos para representar las zonas con más impacto, mientras que los colores fríos representan las zonas menos representativas.

Del mismo modo que en la visualización vía agrupaciones, se suele utilizar en análisis de densidad y en el reconocimiento de patrones en estudios de segmentación de mercado o localización de clientes.

El ejemplo inferior podría indicar la zona donde existe una mayor concentración de puntos (tiendas) en una zona geográfica determinada.

## Ejemplo de mapa de calor

A 10,000-point demo of [Leaflet.heat](#), a tiny and fast Leaflet heatmap plugin.  Star 97

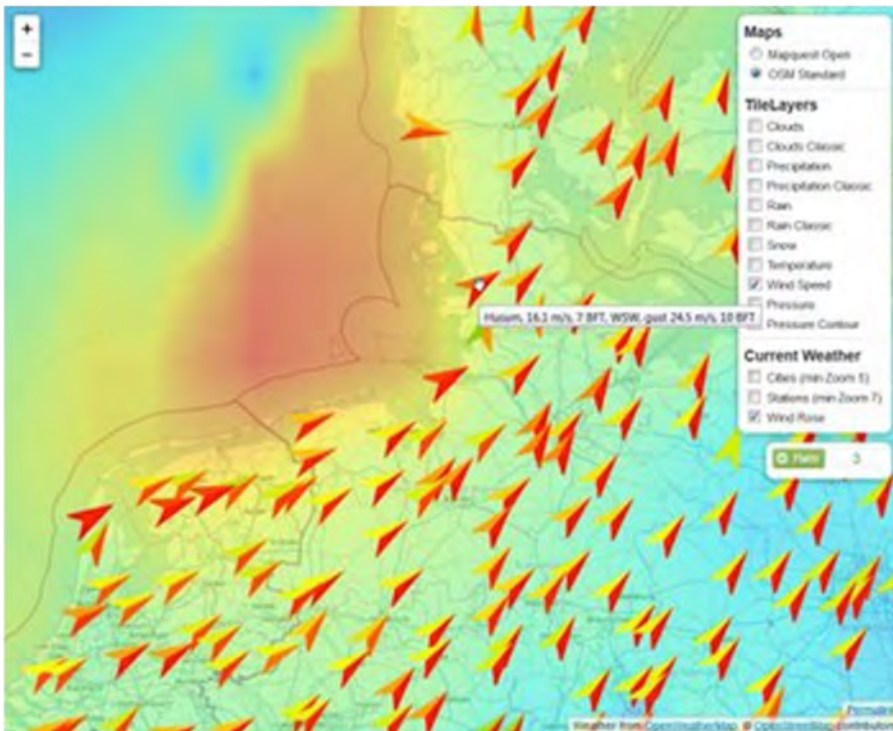


Fuente: <http://leaflet.github.io/Leaflet.heat/demo/>.

### 5.2.3. Mapas de flujos

Un mapa de flujos sirve para ilustrar el movimiento de elementos en un mapa. Para conseguirlo, se unen por las trayectorias pares de localizaciones de la misma entidad. Este tipo de mapa permite definir, entre otros, movimientos migratorios o rutas logísticas.

La velocidad y dirección del viento de la tormenta Christian el 28 de octubre del 2013 en el noroeste de Alemania se sobrepone en un mapa basado en OSM



Fuente: OpenStreetMap.

#### 5.2.4. Colores graduados o coropletas

Es la opción más habitual en la representación de información temática cuantitativa. La clasificación por categorías la encontramos representada mediante el uso de una trama de colores graduados (también llamados coropletas). El símbolo que representa las diferentes categorías es el mismo, mientras que el elemento distintivo es el color.

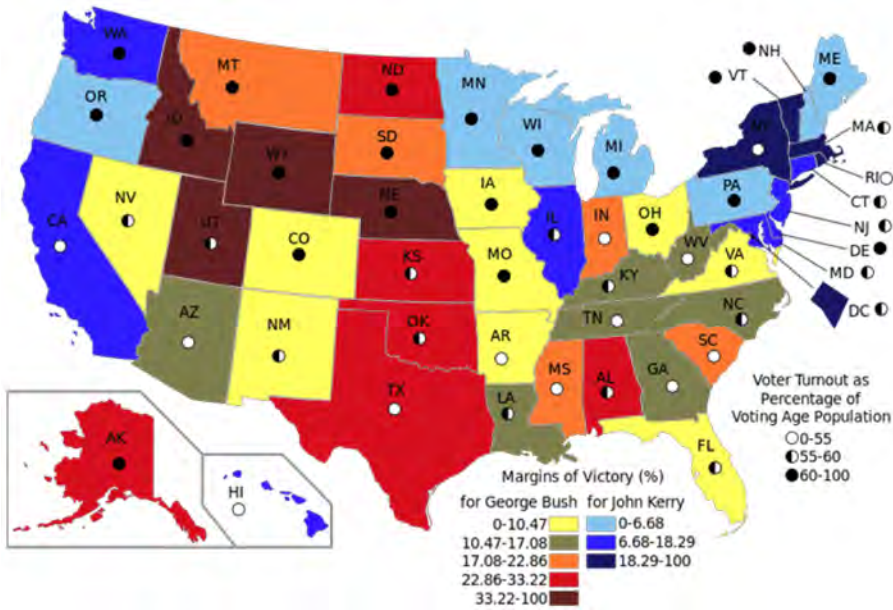
Con el fin de enfatizar el valor que se quiere representar, el rango de colores elegido suele ir de los fríos a los cálidos. En la siguiente figura se pueden ver diferentes opciones de colores graduados.

##### **Nota**

El ojo humano distingue un número limitado de colores y tonalidades. Por ese motivo, es recomendable no utilizar más de siete categorías del mismo color.



Mapa coroplético de las elecciones de Estados Unidos en el 2004

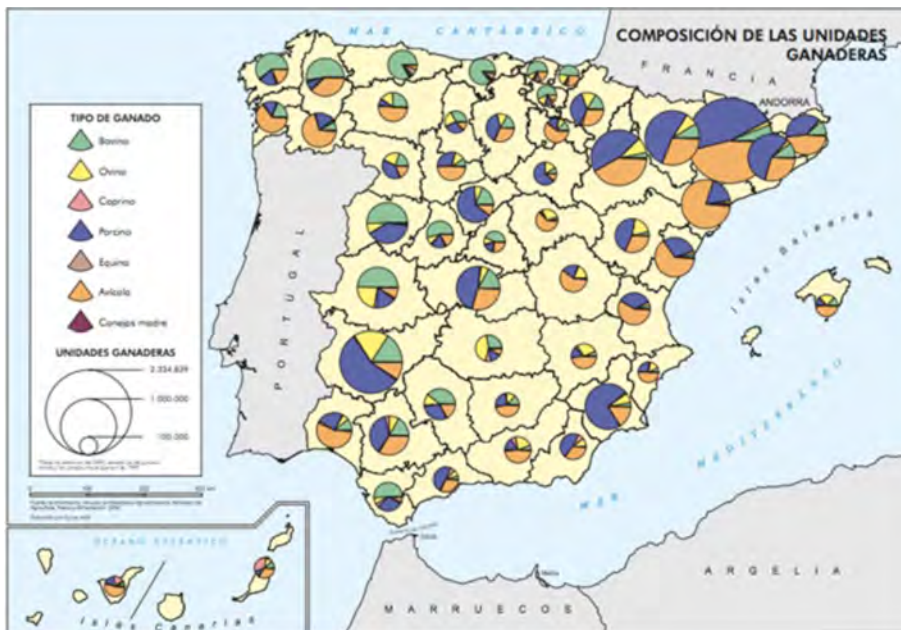


Fuente: Wikipedia.

### 5.2.5. Símbolos graduados

Al igual que la representación de colores graduados, la representación por símbolos graduados también es un método bastante utilizado para representar información temática cuantitativa. En este caso, los elementos utilizan el mismo símbolo, pero no es el color el que determina la magnitud (o valor) de la categoría, sino el tamaño del símbolo.

Mapa de composición de unidades ganaderas por provincia



Fuente: Instituto Geográfico Nacional.



### 5.2.6. Puntos de densidad

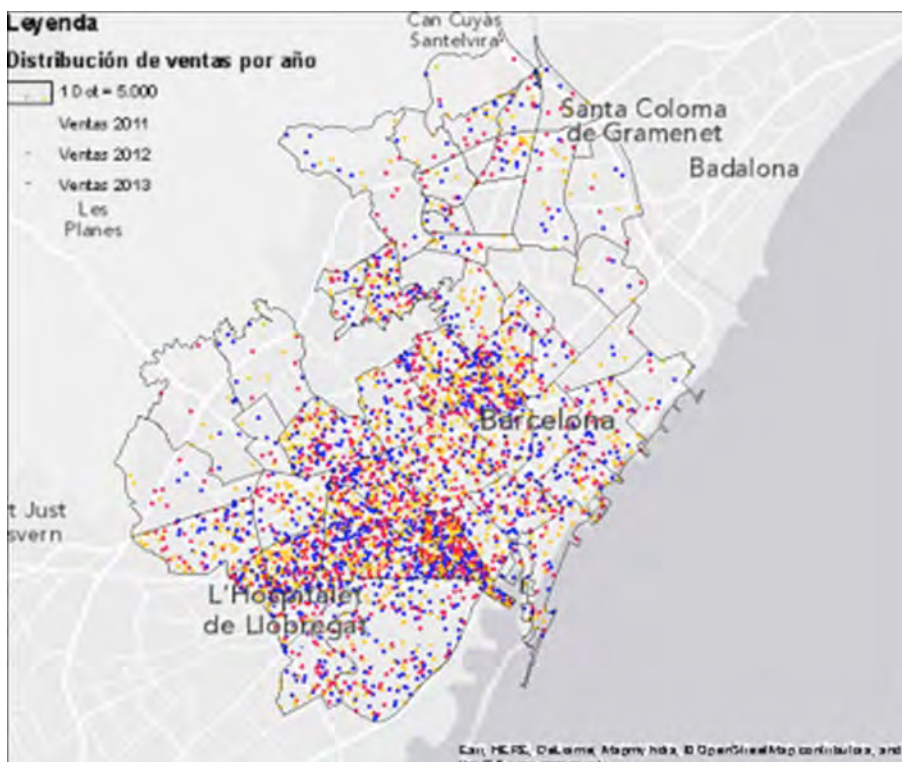
Es similar la representación temática de colores, pero en este caso las variaciones se muestran en la trama en forma de densidad de puntos que rellena la geometría. Este método de representación solo es posible en geometrías de tipo líneas y polígonos.

Este tipo de representación permite visualizar categorías combinadas, por ejemplo: una variable cualitativa (colores por categorías) junto a una variable cuantitativa (valor en densidad de puntos).

En la imagen inferior se muestran las ventas comprendidas entre el 2011 y el 2013 por barrio. Cada punto representa un volumen de ventas de 5.000 €. Los barrios donde existe una mayor densidad de puntos tienen un mayor volumen de ventas.

Los colores de los puntos representan los años de las ventas (amarillo: ventas 2011; rojo: ventas 2012, y azul: ventas 2013).

Mapa de puntos por densidad y año



Fuente: Elaboración propia. Mapa base Esri.

## 6. Análisis de datos

El análisis de datos es un imperativo en los sistemas de BI y en los SIG. Sin la capacidad de interrogar los datos representados, ninguna de estas tecnologías tendría sentido.

Aparte de las operaciones básicas en cualquier análisis de datos, en los SIG se pueden utilizar dos tipos más de operaciones:

- **Operaciones geométricas:** son operaciones cuyo resultado es una extracción geométrica. Como por ejemplo la extracción de los datos contenidos dentro de una capa en función de una geometría concreta. Podría ser el caso de la selección de locales disponibles para la obertura de nuevas sucursales, junto a la capa de densidad de población.
- **Operaciones de combinación:** en las que el resultado proviene de datos de dos o más capas en concreto. Es como sobreponer dos capas de datos. Por ejemplo, la zona de influencia de una determinada sucursal.

Los siguientes apartados tratarán sobre las distintas opciones de análisis. En la primera sección se muestran las operaciones más comunes entre las distintas operaciones de SIG, aplicables a las necesidades de BI; mientras que en la segunda nos centramos en operaciones más específicas del ámbito estricto del negocio.

### 6.1. Operaciones de análisis básicas

Los SIG se utilizan en múltiples ámbitos: generaciones cartográficas, gestión de parcelas, canalizaciones de agua, luz o gas, redes de comunicaciones, gestiones de incidencias, redes de transporte, etc. Para cada uno de estos ámbitos existen herramientas de análisis especializadas. No obstante, existen análisis básicos, comunes entre todos ellos.

Las operaciones básicas sobre datos geográficos se pueden resumir, siguiendo a Pogodzinski y Kos, en ocho puntos:

#### 1) Combinación relacional

Es la típica operación de combinación (*join*) del modelo relacional. La combinación se realiza mediante un campo (atributo) común, ya sea numérico o alfanumérico.

#### 2) Unión espacial

#### Referencia bibliográfica

J. M. Pogodzinski; R. M. Kos (2013). *Economic Development & GIS*. Redlands, California: ESRI Press.

La relación no se establece mediante atributos de datos, sino mediante su geometría, es decir, se tiene en cuenta la disposición de los elementos en el espacio para relacionarlos. Por ejemplo, un supermercado (representado por un punto) contenido en una zona con alta densidad demográfica (polígono). La operación es de combinación y el resultado es la capa original con el contenido, enriquecido con la información de la capa sobre la que combina. El resultado sería el supermercado con la información de población proveniente de la capa de densidad de población.

Consideraremos como entidad de unión el elemento del cual se añade la información, y como entidad destino, aquel elemento que queda enriquecido con la información de la entidad relacionada.

Igual que en el caso de la combinación relacional, cabe definir el marco de relación entre los elementos, que puede ser de distintos tipos:

- **Intersección:** en el caso de que la entidad de unión cruce, contenga o toque la entidad destino.
- **Contener:** en el caso de que la entidad de unión esté contenida dentro de la entidad destino.
- **Concordancia:** en el caso de que las dos entidades sean equivalentes total o parcialmente.
- **Proximidad:** la entidad de unión se relaciona con la entidad destino, siempre que se encuentre dentro de una distancia definida.

### 3) Geocodificación

El proceso de geocodificación consiste en transformar una dirección o punto de interés en coordenadas geográficas (latitud/longitud). El mismo proceso pero en dirección opuesta, es decir, seleccionar un punto en el mapa y obtener la dirección postal se conoce como **geocodificación inversa**.

### 4) Selección por atributos

Consiste en seleccionar aquellos elementos de una capa que satisfacen un determinado criterio de búsqueda. Por ejemplo, todas las carreteras de la categoría "provincial".

### 5) Selección por geometría

Consiste en seleccionar aquellos elementos de una capa que cumplen un determinado criterio de posición geográfica respecto a otra capa. Por ejemplo, las gasolineras situadas dentro de una vía rápida.

## 6) Simbolización temática y simbolización de datos cualitativos

Consiste en simbolizar los elementos diferenciadamente a partir del valor de un campo o atributo de la entidad que representa. Forman parte del método de clasificación y representación cartográfica por categorías descrito en el apartado titulado clasificación por categorías.

## 7) Zonas de influencia o *buffer*

Una zona de influencia o *buffer* es el polígono resultante de aplicar una distancia concreta respecto a las entidades origen. Por ejemplo, un *buffer* lineal de 500 metros alrededor de un establecimiento de la competencia que deseamos contrarrestar será una circunferencia con un radio de 500 metros alrededor de la tienda, que definiría la zona donde localizar a los clientes potenciales.

Zona de influencia de 500 m respecto a la localización de un almacén



Fuente: Elaboración propia; mapa base Esri.

## 8) Intersección

A partir de la relación entre dos capas de información, el proceso de intersección determinará aquellos elementos geométricos que están presentes en ambas capas, conservando los atributos diferenciados que representaban en cada una de ellas. Por ejemplo, dadas dos capas, una de locales disponibles para su alquiler y otra de zonas potenciales de expansión, un proceso de intersección entre ambas capas podría dar lugar a dos posibles resultados, según el orden en el que se ejecute: una capa de locales disponibles que se encuentran en una zona de expansión, o una capa de zonas de expansión donde existen locales disponibles.

La diferencia respecto a la unión espacial es que en este caso las dos capas no tienen una relación estática, sino que está generada en tiempo de análisis.

## 6.2. Operaciones de *geomarketing*

En el ámbito del análisis de negocio basado en datos espaciales, también denominado *geomarketing*, los sistemas de información geográfica describen un conjunto de funcionalidades de análisis orientadas a varios objetivos, como por ejemplo el análisis de clientes o la distribución de zonas comerciales. Este apartado recoge las operaciones más relevantes en este ámbito.

### 6.2.1. Diseño del territorio

Establecer y analizar correctamente las áreas de negocio es un proceso crítico en cualquier estrategia de negocio. A menudo se puede analizar el impacto del negocio en el mercado utilizando los límites administrativos estándar. No obstante, lo habitual es que el negocio defina sus propias áreas, definiendo, por ejemplo, zonas comerciales sobre una ciudad que no coinciden con la distribución de los barrios de esta. Una zona podría comprender medio barrio, dos barrios o simplemente zonas completamente diferentes.

Un correcto estudio de las áreas de negocio permite responder a preguntas como:

- ¿Las áreas definidas son las óptimas?
- ¿Existe alguna opción para mejorar las áreas actuales?
- ¿Existe “canibalización”? Es decir, ¿me hago competencia a mí mismo?
- ¿Cuáles son las zonas comerciales tipo?
- ¿Qué zonas comerciales rinden por debajo de su potencial?

A continuación se presentan algunas de las técnicas comunes para el análisis de áreas de influencia:

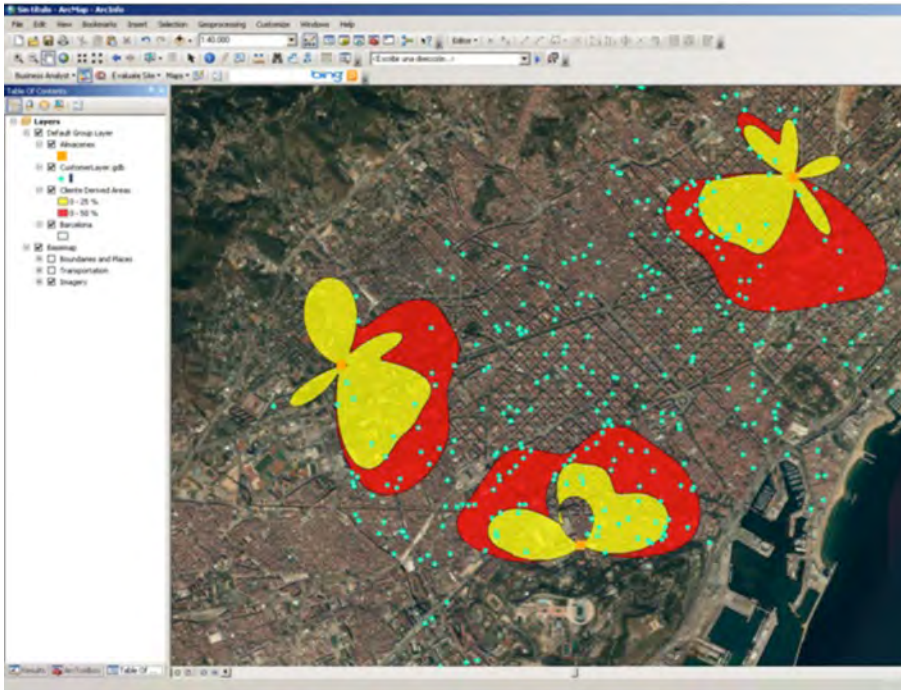
#### Áreas derivadas de cliente

Dada una capa con valores de referencia y una capa origen, las áreas derivadas definen áreas de influencia alrededor del origen teniendo en cuenta la referencia. La capa de referencia puede tener en cuenta un peso. El resultado es un área de influencia.

La siguiente imagen muestra dónde los “almacenes” (cuadrados naranjas) tienen localizados el 25 y 50% de los “clientes” más cercanos.



Localización del 25 y 50% de los clientes alrededor de tres almacenes



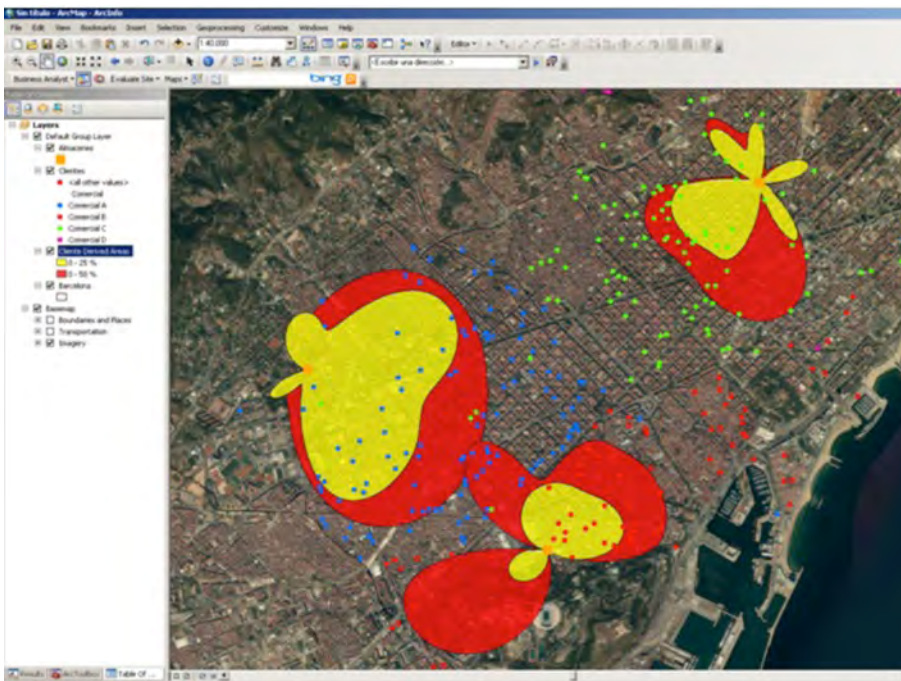
Fuente: Elaboración propia; mapa base Esri.

Si aplicamos el área derivada con pesos (total ventas), podemos ver en la siguiente imagen dónde tienen localizados los “almacenes” el 25 y 50% de su facturación.

Se observa que en este caso el resultado es distinto, pues se considera la ubicación de cada almacén y no la ubicación de sus clientes<sup>3</sup>.

<sup>(3)</sup>Se puede observar cómo hay superposición de zona de influencia clientes entre los almacenes de la parte inferior de la imagen. Estos casos se definen como *canibalización*.

Localización alrededor de los almacenes del 25 y 50% de sus clientes con más facturación



Fuente: Elaboración propia; mapa base Esri.

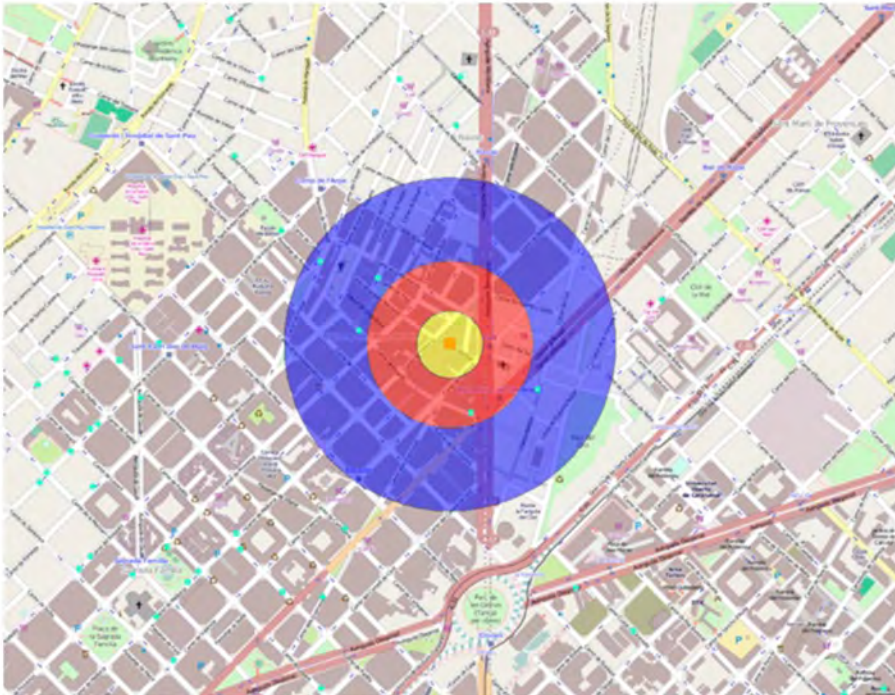


## Anillos simples

Representan una zona de influencia lineal alrededor de una capa origen. Es la técnica de representación más simple y forma parte de la lista de las siete operaciones más comunes descritas en los puntos de análisis básicos y zonas de influencia o *buffers*.

Un posible caso de uso podría ser la localización de comercios a 100, 250 y 500 metros de distancia alrededor de un almacén inicial.

Zona de influencia lineal de 100, 200 y 500 metros alrededor de un almacén



Fuente: Elaboración propia; mapa base OpenStreetMap.

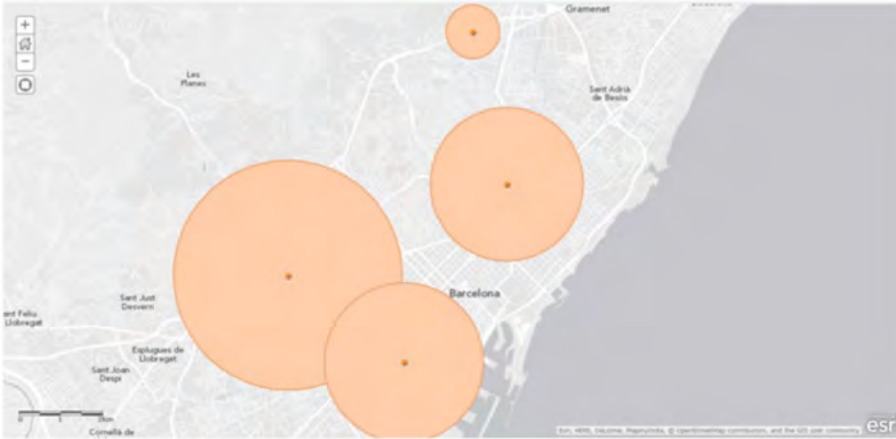
## Anillos guiados

Los anillos guiados son similares al área generada por los anillos simples, pero con la variante de que el radio es ponderado. Un ejemplo de este tipo podrían ser las áreas alrededor de almacenes con un radio equivalente al volumen de ventas.

La siguiente imagen muestra el análisis por anillos guiados de ventas por establecimiento. Cada 100 € del importe de ventas equivale a 1 metro en el radio del anillo en el mapa.

Visualmente, es fácil determinar que el establecimiento que ofrece un volumen de ventas mayor es el situado en el punto más a la izquierda del mapa.

Volúmenes de venta por establecimiento (100 € equivale a un radio de 1 metro)

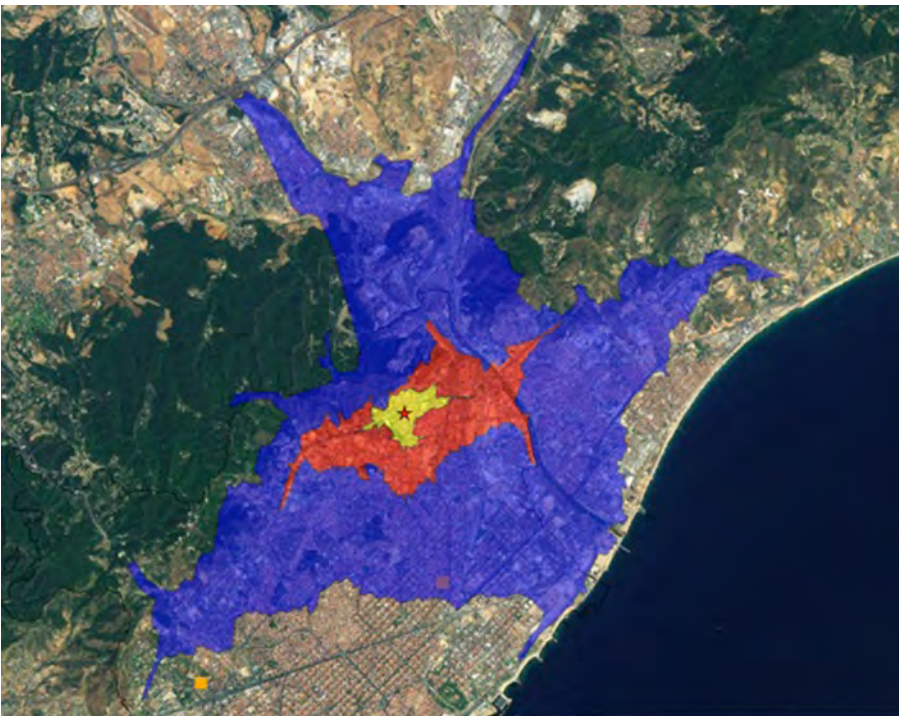


Fuente: Elaboración propia; mapa base Esri.

### Áreas de servicio

Se trata de una evolución de los anillos simples, pero en este caso, en lugar de definir una distancia lineal, se define por tiempos o distancia sobre un callejero. Por ejemplo, zona de influencia de 2, 4 y 8 minutos en coche alrededor de un almacén en Barcelona (imagen siguiente).

Área de servicio de 2, 4 y 8 minutos en coche alrededor de un almacén



Fuente: Elaboración propia; mapa base Esri.

El resultado será un *buffer* irregular, ya que las calles tienen sentido y velocidad. Por tanto, por norma general, no se consigue la misma área en el casco antiguo de una zona urbana, donde posiblemente se puede correr menos, que cerca de una autopista.

## Polígonos de Thiessen o diagrama de Voronoi

La definición de áreas de competitividad (también conocidas como polígonos de Thiessen) tiene en cuenta la proximidad entre los elementos de una misma capa para definir la zona de influencia de cada uno de los puntos. Para ello se unen los puntos y se trazan las mediatrices entre los segmentos resultantes.

La siguiente imagen distribuye la zona de influencia de varios centros veterinarios en la ciudad de Valencia.

Ejemplo de zonas de influencia por clínica veterinaria en Valencia



Fuente: <http://geomarketingvalencia.blogspot.com.es/>.

En un estudio en el que se quiera determinar qué zonas son las ideales para abrir un nuevo centro, se deberían considerar los puntos de confluencia de los polígonos con una superficie mayor. Estos indicarían las zonas con una menor competencia.

La ecuación de probabilidad de Huff es un modelo gravitacional que se define por dos variables: la masa y la fricción. La masa se refiere a los valores que favorecen la localización de puntos de venta (por ejemplo, superficie comercial), mientras que la fricción representa aquellas características que afectan negativamente a la localización (por ejemplo, la distancia respecto a un centro comercial). El doctor David Huff propuso que la utilidad de un establecimiento ( $j$ ) para un comprador ( $i$ ) es:



### Ecuación de probabilidad de Huff

$$P_{ij} = \frac{U_{ij}}{\sum_{k=1}^j U_{ik}} = \frac{S_j^\alpha D_{ij}^\beta}{\sum_{k=1}^j S_k^\alpha D_{ik}^\beta} \quad (1)$$

$P_{ij}$ : es la probabilidad de que el consumidor  $i$  visite el establecimiento  $j$  (o la ciudad  $j$ );  $J$  es el conjunto de establecimientos de la región.

$U_{ij}$ : utilidad de establecimiento  $j$  para el individuo  $i$ .

$S_j$ : superficie comercial (metros cuadrados) del establecimiento  $j$  (o conjunto de establecimientos del municipio  $j$ ).

$D_{ij}$ : distancia entre el consumidor  $i$  y el establecimiento (o ciudad)  $j$ .

$\alpha, \beta$ : parámetros de sensibilidad; suelen definirse  $\alpha = 1, \beta = -2$ .

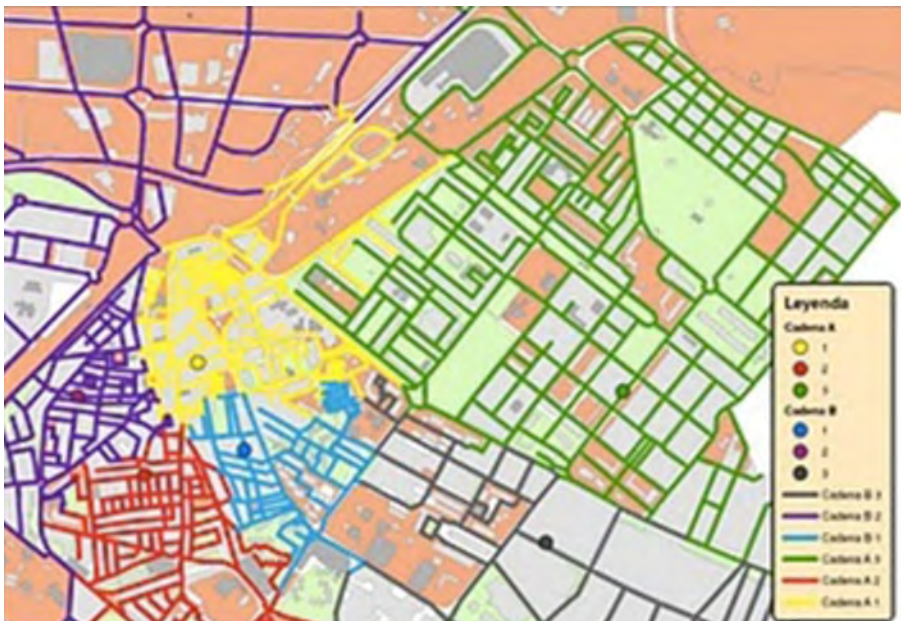
Fuente: <http://www.cartografia.cl/>.

#### Nota

Como se puede observar, para el cálculo de las distancias entre el consumidor ( $i$ ) y el establecimiento ( $j$ ) es necesario tener los datos del entramado de la red de calles.

En un ejemplo desarrollado en la web *Cartografía.cl* se estudia la zona de influencia de dos cadenas A y B con tres establecimientos cada una, y una superficie (en unidades de  $m^2$ ) de 1.600, 1.700 y 2.500, y de 100, 2.500 y 1.800, respectivamente. El resultado lo podemos ver en la siguiente figura:

Zona de influencia de seis centros comerciales utilizando la ecuación de Huff



Fuente: <http://www.cartografia.cl/>.

### 6.2.2. Planificación de mercado

El proceso de expansión comercial es costoso y es un hecho que puede perjudicar gravemente un negocio en caso de fracaso o una mala planificación. Un correcto conocimiento del territorio reduce considerablemente el riesgo de fracaso en estos casos. Los datos geográficos pueden tener un papel muy importante en el estudio de mercado de las empresas, tal y como se comenta a continuación.

Una correcta planificación de mercado para una empresa permite identificar:

- ¿Qué zonas de expansión cumplen los potenciales clientes?
- ¿Cómo se pueden clonar los casos de éxito anteriores de la empresa?
- ¿Cuáles son los elementos clave que rigen el mercado potencial?
- ¿Cómo hacer crecer la probabilidad de éxito disminuyendo el riesgo?
- ¿Cómo crecer sobre un territorio geográficamente desconocido?

A continuación se presentan algunas técnicas que permiten analizar territorios y, por tanto, identificar zonas potencialmente más apropiadas para desarrollar un negocio.

### Prospección de mercado

Consiste en analizar el mercado para identificar al público objetivo y tematizar con categorías sobre los factores clave. Por ejemplo, empresas con una facturación de más de 1 millón de euros que tengan entre 50 y 250 trabajadores. Una vez descubiertas las áreas idóneas, se debe definir el plan de despliegue empresarial.

### Búsqueda de patrones similares

Tomar como referencia el elemento más productivo del negocio e identificar los factores clave es el primer paso para el éxito de este análisis. Por ejemplo, en una cadena de supermercados se analiza cuál de los establecimientos funciona mejor. A continuación se busca el detalle que hace que ese establecimiento destaque sobre los otros, por ejemplo que está en una zona situada cerca de una autopista donde social-demográficamente se sitúa un elevado número de familias jóvenes con un poder adquisitivo medio. Una vez identificado el factor clave, se pueden buscar zonas con características clave similares, es decir, situadas a una distancia equivalente de una carretera principal y con una población demográficamente parecida a la de referencia. De esta manera se intenta reproducir el establecimiento tipo.

### Location-allocation

En realidad se trata de un análisis doble en el que intervienen dos capas de datos: una identifica los puntos de demanda y la segunda las ubicaciones que ofrecen servicio a los puntos de demanda.

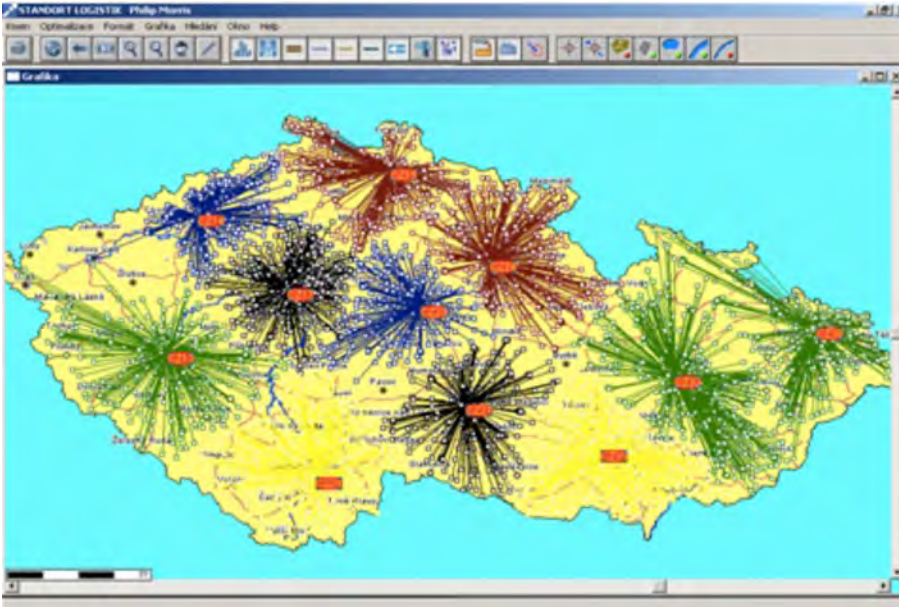
El primer análisis tiene como objetivo encontrar los puntos de demanda principales (clientes potenciales importantes). El segundo análisis determina los centros de distribución (es decir, las ubicaciones) que pueden satisfacer la demanda.

#### Lectura recomendada

Para profundizar este concepto, se recomienda la lectura del capítulo 12 "Estadísticas Espaciales" del libro:  
**V. Olaya** (2012). *Sistemas de Información Geográfica*. Madrid: Bubok.

La siguiente imagen podría representar los puntos de demanda (en forma de población) a menos de 15 minutos en coche que satisface cada centro comercial (puntos de color naranja).

Bolsas de población que cubre cada centro comercial



Fuente: <http://www.logistock.cz/en/modul/storage>.



## 7. Fuentes de datos externas

La información propia de una empresa puede satisfacer la toma de decisiones más comunes. Sin embargo, en algunos casos se necesita una adaptación o ampliación de la información propia para la toma de decisiones geolocalizadas. De hecho, conocer la ubicación de los clientes y de las tiendas de una empresa normalmente permite explicar la razón del comportamiento de los clientes, ya que dicho comportamiento en muchas ocasiones puede depender de factores geográficos externos ligados a su ubicación, como podrían ser localizaciones cercanas a carreteras principales o distancia frente al consumidor tipo.

En el siguiente apartado se citan métodos para tratar fuentes de datos externas que enriquecen el fondo de contenido propio de la empresa y permiten ejecutar preguntas analíticas más potentes. Se cubre tanto la geolocalización de las direcciones postales, el entramado del callejero y los datos estadísticos demográficos o económicos. Debido a la magnitud y a la evolución de las fuentes de datos externas, este capítulo presenta solo algunas opciones disponibles, con el objetivo de invitar al lector a realizar un análisis de mercado en el momento en el que sea necesario para su práctica empresarial.

### 7.1. Normalización para proveer de geolocalización

Los campos de dirección postal de una base de datos pueden ser introducidos en una base de datos sin ningún control. Esta información, en un campo de texto libre, está sujeta a errores cuando no existe un proceso que verifique su integridad y que permita añadir las direcciones postales de manera normalizada. Un proceso de normalización como el propuesto aseguraría la homogeneización de datos. Es decir, que una dirección postal siempre esté escrita de manera completa y del mismo modo, facilitando así que un proceso de geolocalización encuentre las coordenadas de dichas direcciones.

Existen numerosas empresas que ofrecen servicios de normalización de los datos. La siguiente lista muestra algunas de las empresas que ofrecen este servicio basado en la geolocalización:

- **DEYDE** es una empresa privada que provee de un servicio de normalización de direcciones postales y nombres de clientes. Se basa en distintas fuentes de información: ayuntamientos, correos, Instituto Nacional de Estadística y otras cartografías digitales. Su ámbito de actuación es España, Portugal, Italia y Latinoamérica.

- **EGON** una empresa privada que ofrece un servicio de normalización y geocodificación de direcciones postales a nivel mundial, con más de 250 países (incluidos España y América Latina).
- **Correos** también ofrece un servicio de normalización de direcciones postales. Su ámbito es España, Europa, Estados Unidos, Canadá, Australia y Japón.
- La **Junta de Andalucía** ofrece el servicio de normalización (NORDIR) y geocodificación (GEODIR) de direcciones postales libremente, por lo que si el ámbito de actuación empresarial está centrado en esta comunidad autónoma puede ser viable esta opción gratuita. El método es mediante el envío de ficheros KML<sup>4</sup> o CSV.
- **Google** ofrece un servicio de geocodificación y normalización a nivel mundial mediante un servicio web. Permite realizar un total de 2.500 geocodificaciones diarias de manera gratuita<sup>5</sup>.
- **Esri** ofrece desde un servicio de geocodificación de direcciones a nivel mundial mediante un servicio web. Facilita su uso gracias a las numerosas API de desarrollo.

<sup>(4)</sup>El *keyhole markup language* o KML es un estándar geográfico para la representación de datos geográficos en tres dimensiones.

<sup>(5)</sup>Aunque en su contrato de licencia se informa de que estas condiciones pueden cambiar sin previo aviso.

## 7.2. Servicios de callejero

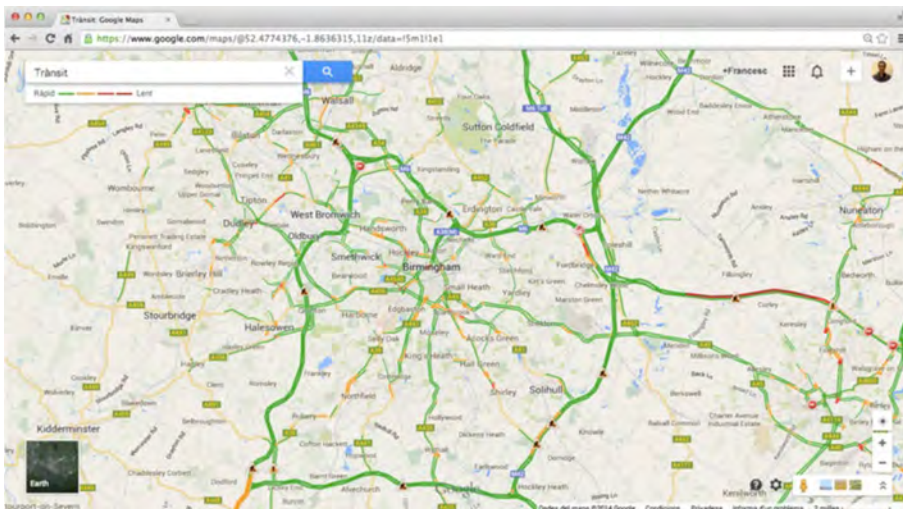
El uso de un modelo de red de transporte (un callejero) es muy habitual en entornos de *geomarketing*: localizar la zona de influencia de un determinado establecimiento, lanzar una campaña de marketing en hogares situados a 5 minutos a pie, determinar la zona de influencia de una cadena de paquetería postal, etc.

A continuación, presentamos algunos proveedores de datos para la generación y/o consumo de servicios de redes de transporte:

- **Nokia/HERE**, anteriormente conocida como NAVTEQ, ofrece un juego de datos para la composición y representación de redes de carreteras, geocodificación y normalización. Incluyen información del tipo: giros prohibidos, barreras físicas, barreras, direcciones únicas, acceso restringido, altura relativa de las carreteras y más de 400 características adicionales. La representación cartográfica es dinámica y en tiempo real, tanto en contenidos como en información sobre el tráfico. Además de los datos de carreteras incluye millones de puntos que indican información de interés general, como la situación de gasolineras, hoteles, tiendas, estadios, hospitales, comisarías de policía, bancos, restaurantes, salas de conciertos, etc.

- **Tele Atlas/TomTom** es una empresa holandesa especializada en información referente a servicios de localización que ofrece datos para la composición y representación de redes de carreteras, geocodificación y normalización. Incluye información detallada del modelo de red de carreteras, así como representación cartográfica en tiempo real del estado del tráfico. Provee servicios avanzados de análisis de redes, como el cálculo de rutas óptimas en una flota de transporte y datos complementarios en forma de distintos puntos de interés distribuidos en 70 categorías (gasolineras, hoteles, tiendas, etc.).
- **Google** ofrece varios servicios web de cálculo de rutas, mediante el ya conocido Google Maps. Por otro lado, Google Transit permite calcular rutas a pie, en bicicleta, en transporte público o en coche. Los cálculos resultantes se integran con el negocio gracias a una API en JavaScript. Junto al servicio de cálculo de rutas, también ofrece la localización y búsqueda de negocios, con cobertura a nivel mundial y análisis del tráfico en tiempo real en más de 600 ciudades de 50 países distintos. En el momento de la captura de la siguiente imagen, en Birmingham, la red de carreteras presenta mayoritariamente fluidez de tráfico, a excepción de los tramos marcados en rojo, que presentan retenciones. Vemos que también proporcionan información sobre tramos de carretera cortados y/o en obras.

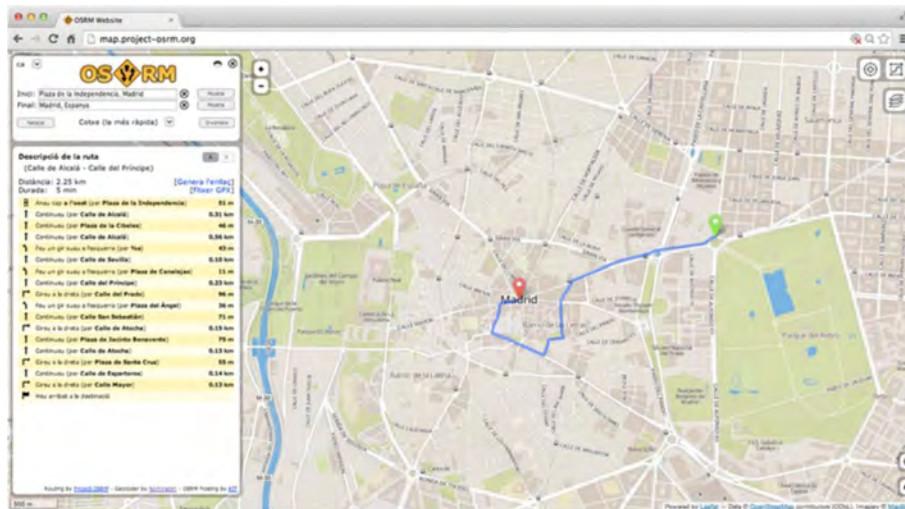
Detalle del servicio de Google en el que se muestra el tráfico en tiempo real en Birmingham



Fuente: Captura de <www.google.com>.

- **OpenStreetMap** es un servicio de mapas abierto y colaborativo. Su cartografía es libre bajo una licencia Open Database License. Existe una comunidad activa de voluntarios que se encargan de revisar, actualizar y expandir el callejero, así como múltiples desarrollos que funcionan sobre este servicio de mapas, como por ejemplo OSRM (*open source routing machine*). OSRM es una API para el cálculo de rutas basado en OpenStreetMap que funciona con una velocidad notable. OpenStreetMap es una buena alternativa, ya que tiene cobertura mundial y el nivel de detalle es bueno, superando en algunos casos al software comercial. En la siguiente imagen se

muestra la resolución del camino más rápido entre dos localizaciones en Madrid utilizando OSRM.



En <map.project-osrm.org> se puede testar la velocidad de ejecución en el cálculo de rutas.

### 7.3. Variables de mercado

A menudo, la toma de decisiones empresariales está íntimamente ligada al enriquecimiento de los datos de la empresa mediante datos provenientes de fuentes de datos externas que contienen información de tipo poblacional, socioeconómica o de mercado.

Existen varias empresas que se encargan de recolectar y tratar datos y que ofrecen servicios de datos geolocalizados. Estos datos pueden responder a distintos temas, como son información empresarial, datos económicos, información relativa a estudios de mercado o variables de población.

A continuación, se describen algunos de los repositorios disponibles en el mercado:

- **AIS** es una empresa consultora especialista en la creación de sistemas de soporte orientados a la toma de decisiones empresariales. Mediante su producto Habits, ofrece un juego de datos con variables social-demográficas cuantificadas, como ingresos y gastos por familia, valor de la vivienda y hábitos de consumo a nivel censal.
- **Data Centric** es una empresa perteneciente al Grupo PDM dedicada al marketing directo y bases de datos en España. En el apartado de datos ofrece un repositorio comercial de tipos de consumidores y negocios por segmento de mercado. Ofrece datos como trabajadores por empresas, categorización por sector, facturación o número de trabajadores.

- **Esri** provee de un servicio SaaS<sup>6</sup> con información de población, negocios y lugares a nivel mundial. Este servicio se puede consumir utilizando un visor web o mediante aplicaciones empresariales de analítica, como pueden ser Excel, Sharepoint, Dynamics, Microstrategy y Cognos, o sistemas CRM como Salesforce. Esri también tiene información de callejero para el análisis de rutas, herramientas de análisis espacial o modelos de elevación.
- **OpenStreetMap**, aparte de ofrecer los datos del entramado de la red de carreteras, también ofrece información sobre localizaciones de puntos de interés, accesibles mediante un servicio web.

<sup>(6)</sup>SaaS o *software as a service* es un modelo de distribución de software como servicio. El cliente se libera del mantenimiento de software y/o datos externalizando el servicio a un servidor remoto al que accede normalmente bajo pago por uso.

## 8. Soluciones de mercado

Los sistemas de SIG y los sistemas de BI provienen de dos disciplinas que están convergiendo debido a las necesidades del mercado, ya sea a partir de la adaptación de la información de negocio a un SIG, ya sea incorporando las capacidades de un SIG dentro de un sistema de BI.

A continuación se describen los diferentes tipos de soluciones que podemos encontrar en el mercado con relación a esta convergencia.

### 8.1. Soluciones SIG

Las soluciones de SIG son capaces de incorporar y tratar información procedente de distintas fuentes de datos empresarial. Ofrecen un gran número de herramientas de análisis, pero al tratarse de programas técnicos requieren una cierta experiencia o conocimiento específico.

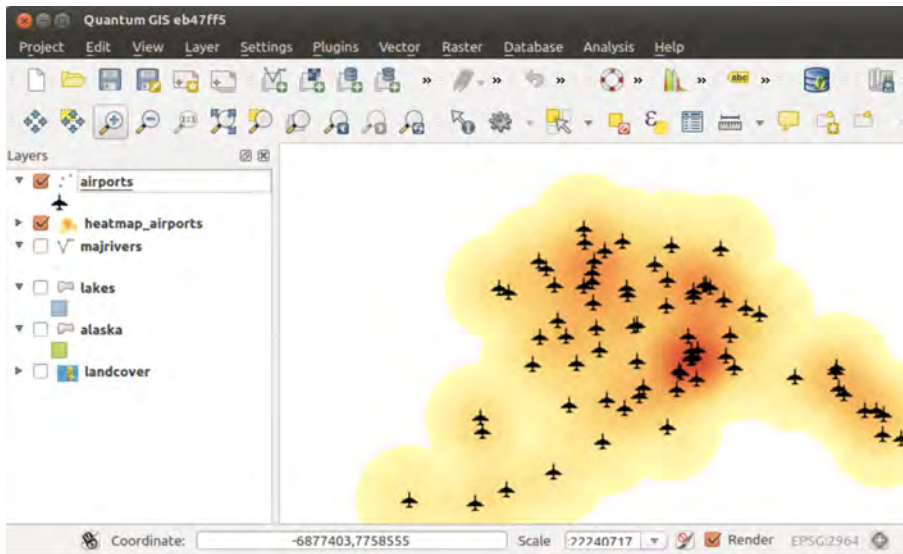
Algunos de los principales proveedores son:

- **Pitney Bowes** con su producto SIG **MapInfo**.
- **Intergraph** con el producto SIG **GeoMedia**.
- **Geoconcept** con el producto específico de *geomarketing* **Geoconcept Geomarketing**.
- **Esri** con la plataforma ArcGIS y su extensión de negocio **Business Analyst**.
- **QGIS** es una solución modular de software libre que permite añadir un componente de *geomarketing* específico.

En la captura siguiente se observa la interfaz de QGIS y un análisis de mapa de calor de la distribución de aeropuertos en Alaska.



### Mapa de calor de concentración de aeropuertos en QGIS



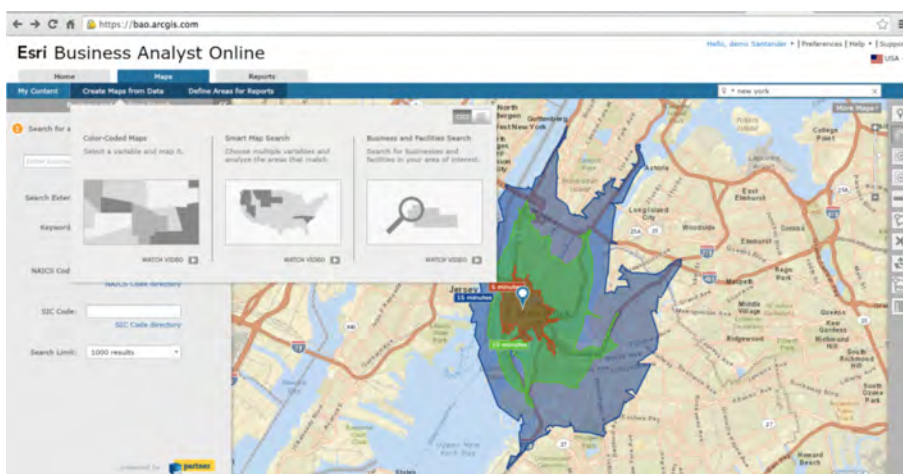
Fuente: [http://docs.qgis.org/2.0/en/docs/user\\_manual/plugins/plugins\\_heatmap.html](http://docs.qgis.org/2.0/en/docs/user_manual/plugins/plugins_heatmap.html).

Todas estas opciones de escritorio ofrecen un largo catálogo de posibilidades de análisis de información geográfica relativa a datos de mercado. Todos ellos son productos de propósito general, a pesar de que tanto Geoconcept como Esri ofrecen adaptaciones orientadas a la explotación de información de negocio que contienen asistentes guiados, análisis avanzados y datos.

Existen variantes a las alternativas de escritorio mostradas, como son las aplicaciones de análisis de negocio web. Business Analyst Online de Esri es un ejemplo de ellas. En estos productos, si bien la funcionalidad no es tan extensa como en las aplicaciones de escritorio, su utilización es mucho más sencilla y en poco tiempo se puede obtener el estudio.

La imagen siguiente ofrece una visión general de la interfaz y del acceso a los asistentes de análisis en Esri Business Analyst Online.

### Interfaz de Esri Business Analyst Online



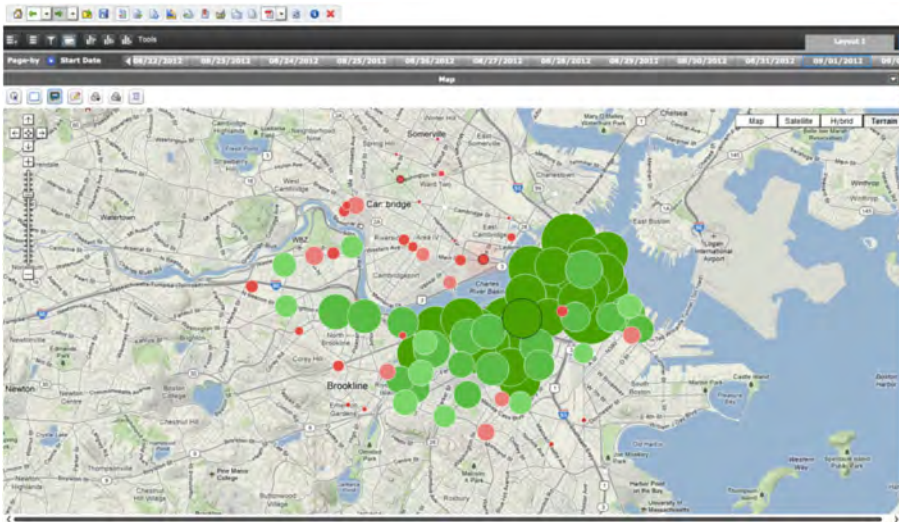
Fuente: Elaboración propia a partir de <http://bao.arcgis.com>.

## 8.2. Soluciones BI

Existen un gran número de productos de BI, por lo que sería imposible describir uno a uno qué componentes geográficos incorporan de serie y con qué productos se pueden extender.

Casi todas las soluciones de BI tienen componentes de serie que permiten la visualización de datos en un mapa, como es el caso de Microstrategy con los componentes Esri Maps y Google Maps (en la imagen inferior) mediante un componente de mapa en un cuadro de mando del BI.

Componente visor de mapas Google Maps de Microstrategy



Fuente: <http://www.chaelchristopher.com/category/microstrategy/>.

Varias empresas ofertan módulos de componentes geográficos que expanden las capacidades de las originales. Por ejemplo:

- El producto **MapIntelligence** de **Integeo** permite incorporar un visor web a múltiples productos BI, como Cognos, Microstrategy, QlikView, SAP BO, etc. Este producto es extensible mediante una API de desarrollo.
- Los productos **Location Analytics** de Esri. Estos productos son individualizados, pero existe un componente para Cognos, Microstrategy, Salesforce, Dynamics, etc. Este producto se integra en la plataforma ArcGIS, por lo que la integración con las fuentes de datos externas especificadas en el apartado anterior es directa.

## Resumen

El presente documento se ha redactado con la intención de aportar un poco de luz a las posibilidades de los SIG aplicados al análisis del sector empresarial (*geographic business intelligence*). Se ha mostrado la tendencia y necesidad de este tipo de sistemas en las soluciones de negocio a partir del uso de los datos geográficos.

Para ello se han introducido los SIG y descrito sus componentes básicos y fundamentos, como es la naturaleza de los datos geográficos, la proyección de la información en un mapa, los modelos de datos existentes, los métodos de reclasificación de estos o las relaciones entre capas de información y recursos para la visualización temporal.

En un contexto de negocio se han descrito las técnicas más utilizadas para la comunicación del contenido geográfico, así como los recursos que se han de utilizar según el modelo y origen del dato para una mejor visualización. Y más allá de la visualización se han enumerado algunas operaciones de análisis basadas en la localización que permiten extraer nueva información, nuevas variables sobre las que valorar el negocio. Finalmente, se han recopilado algunos de los principales recursos que enriquecen la propia base empresarial a partir de fuentes de datos externas.

Este documento no se centra en ninguna solución existente en el mercado, pretende ayudar al lector en la comprensión de las herramientas (y carencias) que describen los distintos productos *geographic business intelligence* existentes, ya sean aquellos que importan los datos de la empresa en un entorno de SIG, ya sean aquellos componentes geográficos que se integran dentro de las propias soluciones de BI.

Se trata de un inicio en el camino del *geographic business intelligence* y una invitación a la continua exploración.



## Glosario

**análisis (espacial)** *m* Análisis que se basa en los aspectos principales de las herramientas lógicas que forman un sistema de información geográfica, como por ejemplo superposición de capas de información, construcción de modelos tridimensionales del terreno y operaciones que permiten la transformación de escalas y coordenadas.

**atributo (geográfico)** *m* Información sobre una entidad geográfica en un sistema de información geográfica. La información se almacena normalmente en una tabla que está enlazada a la entidad por un único identificador.

**callejero** *m* Guía o mapa que muestra las carreteras y las calles de un distrito o de una ciudad entera. Los callejeros son una gran herramienta para navegar por las ciudades, pueblos o comunidades. Originariamente estos estaban disponibles en formato impreso como una guía urbana con sus mapas o planos, actualmente sin embargo es más común su formato digital consultable a través de internet o de dispositivos móviles con conexión GPRS.

**capa** *f* Representación visual de la información geográfica en un mapa digital. Concretamente, una capa es una porción o estrato de la realidad geográfica de un área particular. Por ejemplo, en un mapa de carreteras, los parques nacionales, las fronteras políticas y los ríos son representados en diferentes capas.

**dato alfanumérico** *m* Información nominal y numérica. Los SIG están compuestos de una base de datos con información alfanumérica.

**dato geográfico** *m* Dato con información geográfica.

**geocodificación** *f* Proceso de asignar coordenadas geográficas (por ejemplo, latitud-longitud) a puntos del mapa (direcciones, puntos de interés, etc.). Las coordenadas geográficas producidas pueden luego ser usadas para localizar el punto del mapa en un sistema de información geográfica.

**geolocalización/georreferenciación** *f* Identificación de la ubicación geográfica en el mundo real de un objeto. La geolocalización está estrechamente relacionada con el uso de sistemas de posicionamiento (por ejemplo, GPS), pero puede distinguirse de estos por un mayor énfasis en determinar una posición significativa (por ejemplo, una dirección de calle) y no solo un conjunto de coordenadas geográficas.

**geodesia** *f* Rama de las geociencias y la ingeniería. Engloba las técnicas para la representación de la superficie de la Tierra, con sus formas naturales y artificiales.

**geofencing/geofence** *m* Tecnología que define un límite virtual alrededor de un área geográfica en el mundo real. De este modo, se establece un radio de interés que puede desencadenar una acción a partir de un dispositivo móvil. *Geofencing* permite generar alertas automáticas que se generan en función de las coordenadas definidas en un área geográfica. Un ejemplo sencillo de esto podría ser un correo electrónico o un mensaje de texto (por ejemplo, una oferta comercial) que se activa automáticamente y se envía al teléfono móvil del usuario cuando este se encuentra en un radio de 10 metros de un centro comercial.

**modelado** *m* Proceso que permite generar una representación (modelo) idealizada y simplificada de la realidad.

**POI (punto de interés)** *m* Objeto representado por un punto en un mapa (por ejemplo, un bar o una farola). La característica de "interés" es bastante subjetiva por lo que hace el término un tanto impreciso. Aun así, se trata de un concepto ampliamente reconocido por los usuarios de los sistemas de geoposicionamiento, que a menudo se presentan con opciones para mostrar u ocultar puntos de interés.

**proyección cartográfica** *f* Sistema que convierte la superficie esférica de la Tierra en una superficie plana.

**raster** *adj* Estructura de datos representados en una matriz de celdas regulares. Cada una ellas almacena información geográfica.

**sistema de coordenadas** *m* Sistema que utiliza un conjunto de valores para determinar de manera única la posición de cualquier punto en el espacio con respecto a un punto de referencia.

**topología** *f* Disciplina matemática que estudia las propiedades de proximidad, conectividad, compacidad, textura, metricidad, etc., de las entidades.

**vectorial** *adj* Relativo a estructuras de datos que representan la realidad mediante puntos, líneas y polígonos.



## Bibliografía

- Bouillé, F.** (1978). "Structuring cartographic data and spatial processes with the hypergraph-based data structure". En: Geoffrey Dutton (ed.). *First International Symposium on Topological Data Structures for GIS*. Cambridge, Massachusetts: Laboratory for Computer Graphics and Spatial Analysis, Harvard University.
- Burns, T.** (2006). *Finding Your Customers: GIS for Retail Management*. Redlands, California: ESRI Press.
- Burrough, P. A.; McDonnell, R. A.** (1998). *Principles of geographic information systems for land resources assessment*. Oxford: Clarendon.
- Comas, D.; Ruiz, E.** (1993). *Fundamentos de los Sistemas de Información Geográfica*. Barcelona: Ariel.
- Church, R. L.; Murray, A. T.** (2008). *Business Site Selection, Location Analysis and GIS*. Nueva York: Wiley.
- Davenport, T. H.; Harris, J. G.** (2007). *Competing on Analytics: The New Science of Winning*. Nueva York: Harvard Business Review Press.
- Dueker, K. J.; Kjerne, D.** (1989). *Multipurpose cadastre: Terms and definitions*. Falls Church, Virginia: American Society for Photography and Remote Sensing and American Congress on Surveying and Mapping.
- Pourabbas, E.** (2014). *Geographical Information Systems: Trends and Technologies*. Boca Ratón, Florida: CRC Press.
- Longley, P. A.; Goodchild, M. F.; Maguire, D. J. y otros** (2005). *Geographic Information Systems and Science*. Chichester (Reino Unido): John Wiley & Sons.
- Madden, M.** (ed.) (2009). *Manual of Geographical Information Systems*. Falls Church, Virginia: American Society for Photography and Remote Sensing.
- Maguire, D.; Goodchild, M.; Rhind, D.** (eds.) (1991). *Geographical Information Systems: Principles and Applications*. Avon: Longman Scientific and Technical.
- Maguire, D; Kouyoumjian, V; Smith, R.** (2008). *The Business Benefits of GIS: An ROI Approach*. Redlands, California: ESRI Press.
- Miller, F. L.** (2007). *GIS Tutorial for Marketing Spiral-bound*. Redlands, California: ESRI Press.
- Mitchel, A.** (1999). *GIS Analysis, Volume 1: Geographic Patterns & Relationships*. Redlands, California: ESRI Press.
- Olaya, V.** (2011). *Libro Libre SIG*.
- Perez, A. (coord.); Botella, A.; Muñoz, A.; Olivella, R.; Olmedillas, J. C.; Rodríguez, J.** (2011). *Introducción a los sistemas de información geográfica y geotelemática*. Barcelona: Editorial UOC.
- Pogodzinski, J. M.; Kos, R. M.** (2013). *Economic Development & GIS*. Redlands, California: ESRI Press.
- Schietzelt, T. H.; Densham, P. J.** (2003). "Location-allocation in GIS". En: Longley, P. A.; Batty, M. (eds.). *Advanced Spatial Analysis: The CASA Book of GIS* (págs. 345-368). Redlands, California: ESRI Press.
- Tomilson, R.** (2003). *Thinking about GIS. Geographic Information System Planning for managers*. Redlands, California: ESRI Press.

