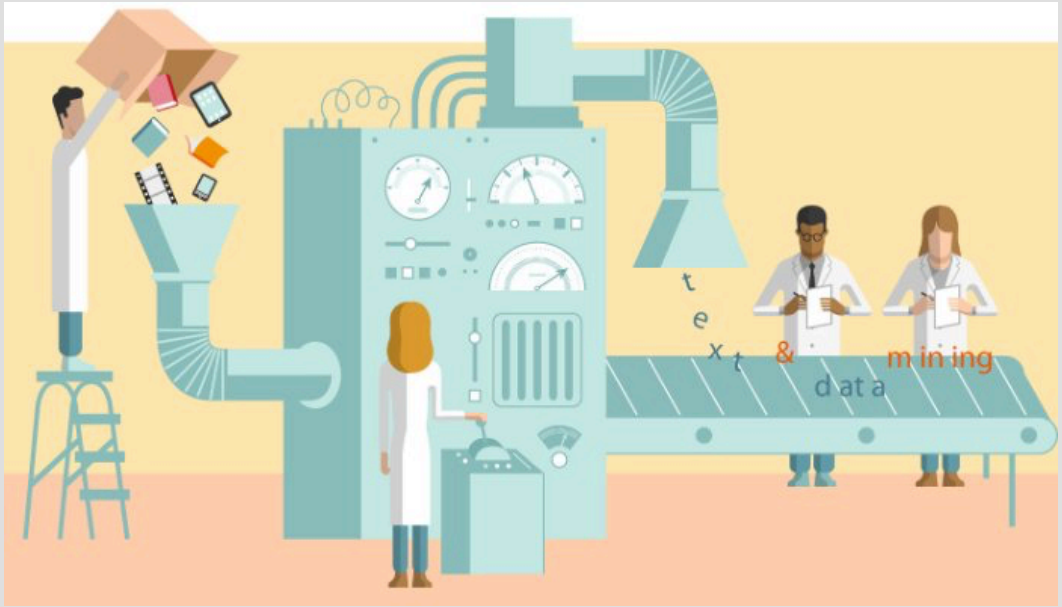


DigitalHouse > Coding School

DATA SCIENCE

MÓDULO 2

GeoPandas

GeoPandas

Datos Geoespaciales

- Una base de datos geoespacial (GIS) es una base de datos referenciada de alguna manera a una ubicación sobre la tierra.
- Junto con estos datos suele ser estar datos conocidos como datos de atributos. Los datos de atributos generalmente se definen como información adicional, que luego puede vincularse a datos espaciales.
- Permiten operaciones no habituales en otros tipos de datos

Datos Geoespaciales

- Los datos GIS pueden separarse en dos categorías: datos referenciados espacialmente representados por vectores y formas ráster (incluidas imágenes) y tablas de atributos que se representan en formato tabular.
- Dentro del grupo de datos de referencia espacial, los datos GIS se pueden clasificar en dos tipos diferentes: vector y ráster. La mayoría de las aplicaciones de software GIS se centran principalmente en el uso y la manipulación de geodatabases vectoriales con componentes adicionales para trabajar con geodata basadas en ráster.

Datos Vectoriales

- Tienen coordenadas y relaciones entre esas coordenadas. .
- Los datos vectoriales se dividen en tres tipos: polígono, línea (o arco) y datos de puntos.
- Los polígonos se utilizan para representar áreas tales como el límite de una ciudad (en un mapa a gran escala), un lago o un bosque. Las características del polígono son bidimensionales y, por lo tanto, se pueden usar para medir el área y el perímetro de una característica geográfica.
- Las características del polígono se distinguen más comúnmente utilizando una simbología de mapeo temático (esquemas de color), patrones, o en el caso de la gradación numérica, se podría usar un esquema de gradación de color.

Datos Vectoriales

OGC Simple Features

Point, LineString, Polygon

... and MultiPoint, MultiLineString, MultiPolygon, GeometryCollection

Datos Vectoriales

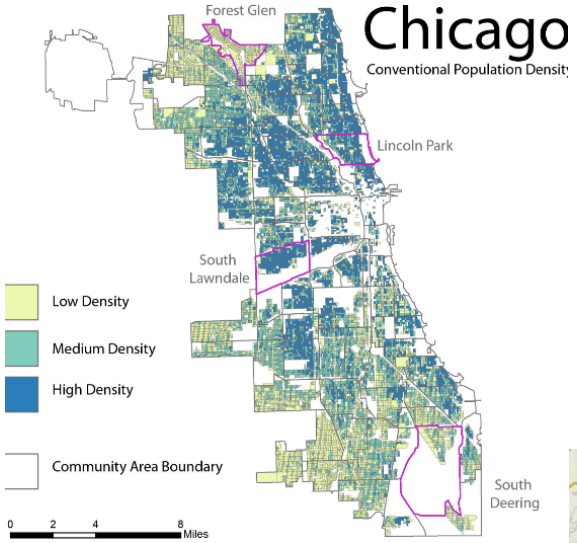
- Los datos de línea (o arco) se utilizan para representar características lineales. Ejemplos comunes serían ríos, senderos y calles. Las entidades de línea solo tienen una dimensión y, por lo tanto, solo se pueden usar para medir la longitud. Las características de la línea tienen un punto inicial y un punto final.
- Ejemplos comunes serían las líneas centrales de carreteras y la hidrología. La simbología más comúnmente utilizada para distinguir las características del arco entre sí son los tipos de línea (líneas continuas versus líneas punteadas) y las combinaciones que utilizan colores y grosores de línea.

Datos Vectoriales

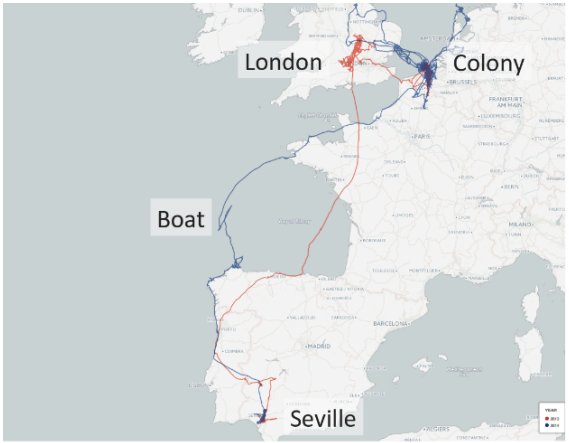
- Los datos de puntos se utilizan con mayor frecuencia para representar entidades no adyacentes y para representar puntos de datos discretos.
- Los puntos tienen cero dimensiones por lo tanto no puede medir ni la longitud ni el área con este conjunto de datos. Ejemplos serían puntos de interés (locales bailables).
- Las características del punto también se usan para representar puntos abstractos. Por ejemplo, las ubicaciones de puntos podrían representar ubicaciones de ciudades o nombres de lugares.

Datos Vectoriales Polígono, línea y punto

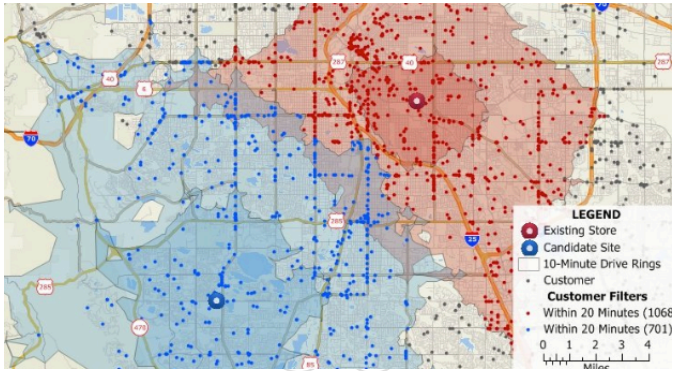
Polygon Features



Line



Point



Datos Vectoriales

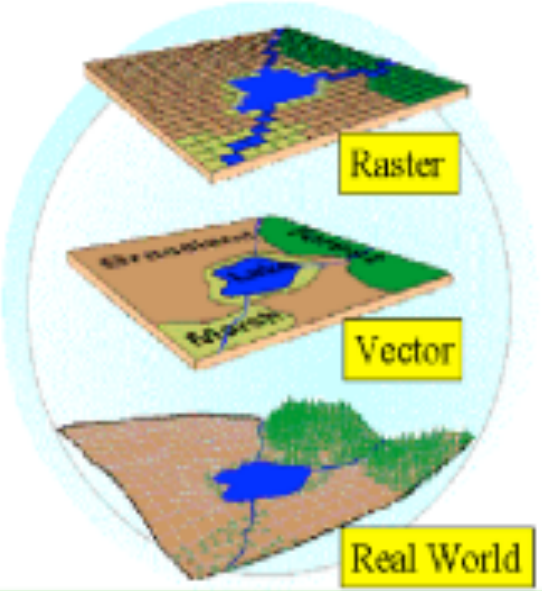
- Las estructuras tanto de línea como de punto representan datos de polígono en una escala mucho más pequeña. Ayudan a reducir el desorden al simplificar las ubicaciones de datos.
- A medida que se hace un zoom in de los features la ubicación del punto de una escuela se representa de forma más realista mediante una serie de huellas de construcción que muestran la ubicación física del campus. Las características de línea de un archivo de línea central de calle solo representan la ubicación física de la calle.
- Si se necesita un mayor grado de resolución espacial, se usará un archivo de ancho de curva de la calle para mostrar el ancho de la carretera y cualquier característica como medianas y derechos de paso (o aceras).

Raster vs vector data

Datos Ruster

- Los ruster data (también conocidos como grid data) representan pixeles.
- Representan el cuarto tipo de característica: superficies.
- Los datos ráster están basados en celdas y esta categoría de datos también incluye imágenes aéreas y de satélite. Hay dos tipos de datos de ráster: continuo y discreto. Un ejemplo de datos ráster discretos es la densidad de población. Los ejemplos de datos continuos son mediciones de temperatura y elevación. También hay tres tipos de datasets ráster: datos temáticos, datos espectrales e imágenes.

Datos Ruster



Vector Format

Raster Format
200 foot pixel size

100 Acre Pond

Datos Ruster

- Lo que resulta del efecto de convertir la información de ubicación de datos espaciales en un formato de ráster basado en celdas se denomina paso por escalera. El nombre deriva de la imagen de eso exactamente, las celdas cuadradas a lo largo de los bordes de diferentes tipos de valores se ven como una escalera vista desde un lado.
- A diferencia de los datos vectoriales, los datos ruster están formados por cada celda que recibe el valor de la característica que domina la celda. El aspecto de paso de escalera proviene de la transición de las celdas de un valor a otro.

Formatos de datos geoespaciales

Lista de formatos

The Ultimate List of GIS Formats – Geospatial File Extensions

GIS Formats

Ecosistema de herramientas de software

Open source geospatial software

GDAL/OGR

- Lee y escribe datasets Raster (GDAL) y Vector (OGR)
- Más de 200 (principalmente) formatos y protocolos geoespaciales.

GEOS Geometry Engine Open Source

DigitalHouse > Coding School

- Puerto C / C ++ de un subconjunto de Java Topology Suite (JTS)
- Biblioteca de geometría C ++ geoespacial más ampliamente utilizada
- Implementa objetos de geometría (características simples), funciones de predicado espacial y operaciones espaciales
- Utilizado under the hood por muchas aplicaciones (QGIS, PostGIS, MapServer, GRASS, GeoDjango, ...).

GEOS Geometry Engine Open Source

Paquetes geoespaciales en Python

DigitalHouse > Coding School

- Interfaces para bibliotecas ampliamente utilizadas:
  - Python binding a GDAL / OGR (from osgeo import gdal, ogr)
  - Interfaz de Python para PROJ.4: pyproj
  - Pythonic binding a GDAL / OGR:
  - GDAL: rasterio
  - OGR: fiona
  - Paquete de Python basado en GEOS: shapely

Ecosistema

- geoplot (visualización geoespacial de alto nivel), cartopy (biblioteca cartográfica projection aware)
- folium (mapas Leaflet.js)
- OSMnx (python para redes de calles)
- PySAL (Biblioteca de análisis espacial de Python)
- rasterio (trabajando con datos geoespaciales de tipo raster)

Shapely

- Paquete Python para la manipulación y análisis de objetos geométricos
- Pythonic interface para GEOS

```
>>> from shapely.geometry import Point, LineString, Polygon

>>> point = Point(1, 1)
>>> line = LineString([(0, 0), (1, 2), (2, 2)])
>>> poly = line.buffer(1)
```

Shapely

- Paquete Python para la manipulación y análisis de objetos geométricos
- Pythonic interface para GEOS

```
>>> from shapely.geometry import Point, LineString, Polygon

>>> point = Point(1, 1)
>>> line = LineString([(0, 0), (1, 2), (2, 2)])
>>> poly = line.buffer(1)
```

GeoPandas

- Facilita el trabajo con datos geoespaciales en Python
- Extiende la biblioteca de análisis de datos de pandas para trabajar con objetos geográficos y operaciones espaciales
- Combina el poder de todo el ecosistema de (geo) herramientas (pandas, geos, shapely, gdal, fiona, pyproj, rtree, ...)

GeoPandas

- GeoPandas puede ser lento
- GeoPandas almacena custom objects de Python en arrays
- Para las operaciones, itera a través de esos objetos
- Geometry Array como nuevo desarrollo para resolver esto

GeoPandas

- GeoPandas implementa dos principales estructuras de datos GeoSeries yGeoDataFrame.
    - Estas son subclases de la Series y DataFrame de pandas respectivamente.
- Una GeoSeries es esencialmente un array donde cada entrada es un conjunto de formas que corresponde a una observación.
- Una entrada puede consistir en una sola forma (como un solo polígono) o múltiples formas que deben considerarse como una sola observación (como los muchos polígonos que componen la Capital Federal o un país como Argentina).

GeoPandas

- Geopandas tiene tres clases básicas de objetos geométricos (que en realidad son objetos bien formados):
  - Puntos / Múltiples Puntos
  - Líneas / líneas múltiples
  - Polígonos / polígonos múltiples
- No es necesario que todas las entradas en una GeoSeries sean del mismo tipo geométrico, aunque algunas operaciones de exportación fallan si este no es el caso.

GeoPandas

- La clase GeoSeries implementa casi todos los atributos y métodos de objetos Shapely. Cuando se aplican a una GeoSeries, se aplicarán de forma automática a todas las geometrías de la serie.
- Las operaciones binarias se pueden aplicar entre dos GeoSeries, en cuyo caso la operación se lleva a cabo por elementos. Las dos series se alinearán mediante índices coincidentes.
- Las operaciones binarias también se pueden aplicar a una única geometría, en cuyo caso la operación se lleva a cabo para cada elemento de la serie con esa geometría. En cualquier caso, se devolverá una Serie o una GeoSeries, según corresponda.

Atributos y métodos de una geoserie

Attributes

- `area` : shape area (units of projection – see projections)
- `bounds` : tuple of max and min coordinates on each axis for each shape
- `total_bounds` : tuple of max and min coordinates on each axis for entire GeoSeries
- `geom_type` : type of geometry.
- `is_valid` : tests if coordinates make a shape that is reasonable geometric shape (according to

Basic Methods

- `distance(other)` : returns `Series` with minimum distance from each entry to `other`
- `centroid` : returns `GeoSeries` of centroids
- `representative_point()` : returns `GeoSeries` of points that are guaranteed to be within each geometry. It does **NOT** return centroids.
- `to_crs()` : change coordinate reference system. See projections
- `plot()` : plot `GeoSeries`. See mapping.

GeoPandas

- Un GeoDataFrame es una estructura de datos tabulares que contiene una GeoSeries.
- La propiedad más importante de un GeoDataFrame es que siempre tiene una columna GeoSeries que tiene un status especial.
- Esta GeoSeries se conoce como "geometría" de GeoDataFrame. Cuando se aplica un método espacial a un GeoDataFrame (o se llama un atributo espacial como área), estos comandos siempre actuarán en la columna "geometría".

GeoPandas

- Se puede acceder a la columna "geometría", sin importar su nombre, a través del atributo de geometría (gdf.geometry), y se puede encontrar el nombre de la columna de geometría escribiendo gdf.geometry.name.
- Un GeoDataFrame también puede contener otras columnas con objetos geométricos (con formas), pero solo una columna puede ser la geometría activa a la vez.
- Para cambiar qué columna es la columna de geometría activa, use el método set_geometry.

- Un GeoDataFrame realiza un seguimiento de la columna activa por nombre, por lo que si cambia el nombre de la columna de geometría activa, también debe restablecer la geometría:

```
gdf = gdf.rename(columns={'old_name': 'new_name'}).set_geometry('new_name')
```

- Cualquiera de los calls a atributos o métodos descritos para una GeoSeries funcionará en un GeoDataFrame; de hecho, solo se aplican a la GeoSeries "geometría". Sin embargo, GeoDataFrames también tiene algunos métodos adicionales para entrada y salida

- De manera algo confusa, de forma predeterminada, cuando utiliza el comando read_file, la columna que contiene objetos espaciales del archivo se denomina "geometría" de forma predeterminada y se configurará como la columna de geometría activa.
- Sin embargo, a pesar de utilizar el mismo término para el nombre de la columna y el nombre del atributo especial que realiza un seguimiento de la columna activa, son distintos.

GeoPandas

- Puede cambiar fácilmente la columna de geometría activa a una GeoSeries diferente con el comando set_geometry.
- Además, gdf.geometry siempre devolverá la columna de geometría activa, no la columna denominada geometría.
- Si desea llamar a una columna llamada "geometría", y una columna diferente es la columna de geometría activa, use gdf ['geometry'], no gdf.geometry.

Práctica Guiada GeoPandas

Lab