

Workshop CIENFI: Introducción al Análisis de Datos Espaciales en R

Eduard F. Martínez González, Ph.D.

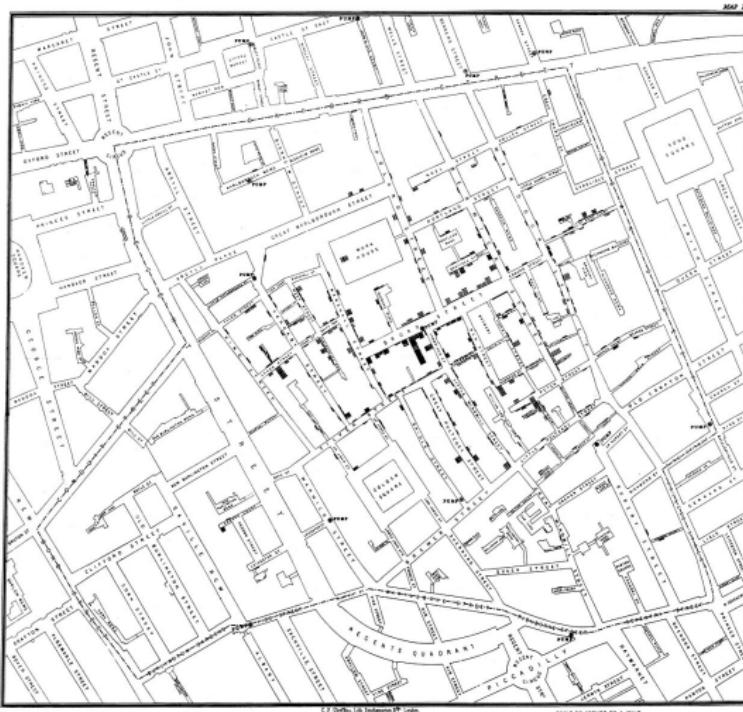
Centro de Investigación en Economía y Finanzas (CIENFI)
Universidad Icesi

Marzo 2025

Motivación: John Snow y el Mapa del Córula

- **Brote de cólera en el Soho de Londres (1854):**
 - ▶ Aumento inusual de muertes concentradas en un área específica.
 - ▶ Causa atribuida inicialmente a “miasmas” (aire contaminado).
- **John Snow plantea una hipótesis alternativa:**
 - ▶ El cólera no se transmite por el aire, sino por el agua.
- **Recolección de datos y análisis:**
 - ▶ Registra ubicaciones de muertes por cólera.
 - ▶ Mapea los pozos de agua del vecindario.
- **Visualización espacial y evidencia empírica que respalda su hipótesis:**
 - ▶ Identifica un patrón claro alrededor de la bomba de agua en Broad Street.
 - ▶ Se retira la manivela de la bomba; los casos disminuyen.

Motivación: John Snow y el Mapa del Córula



Nota: Tomado de <https://es.wikipedia.org>

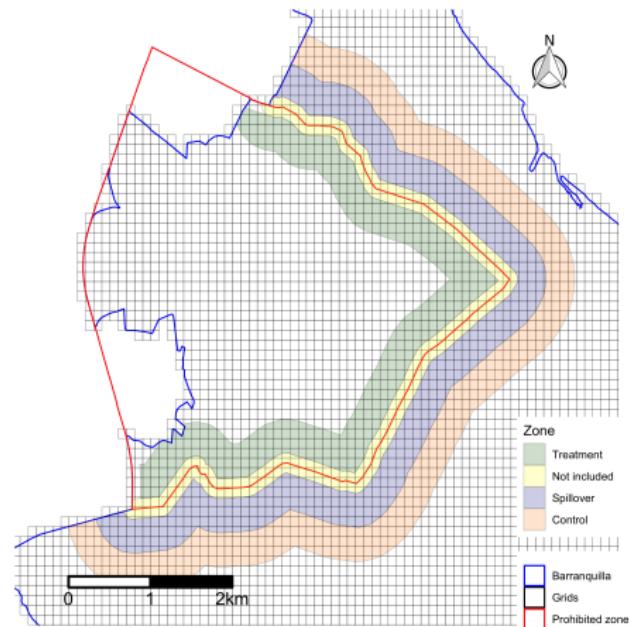
1 Motivación

2 Espacio, Datos y Causalidad: Evidencia Reciente

3 Introducción a Datos Espaciales

4 Introducción al Ecosistema de Datos Espaciales en R

¿Reducen las restricciones a la movilidad en motocicleta los niveles de crimen en áreas urbanas?



Fuente: Martínez-González et al. (2022)

Restrictions on Urban Mobility and Crime: Evidence from Motorcycle Bans in Colombia

- **Datos:**

- ▶ Datos geo-referenciados de crímenes (100m x 100m).
- ▶ Restricciones aplicadas en distintas ciudades y momentos.

- **Estrategia empírica:**

- ▶ *Spatial Regression Discontinuity Design (SRDD)*.
- ▶ Diferencias-en-diferencias con zonas tratadas, de spillover y de control.

- **Resultados principales:**

- ▶ En la mayoría de los casos, las restricciones no reducen el crimen.
- ▶ Donde hay efectos, se observa desplazamiento espacial del crimen.
- ▶ Mejor orientar esfuerzos a políticas focalizadas y con alta precisión.

Midiendo el crecimiento económico desde el espacio

Henderson, Storeygard y Weil (2012) – American Economic Review

Measuring Economic Growth from Outer Space[†]

By J. VERNON HENDERSON, ADAM STOREYGARD, AND DAVID N. WEIL*

We develop a statistical framework to use satellite data on night lights to augment official income growth measures. For countries with poor national income accounts, the optimal estimate of growth is a composite with roughly equal weights on conventionally measured growth and growth predicted from lights. Our estimates differ from official data by up to three percentage points annually. Using lights, empirical analyses of growth need no longer use countries as the unit of analysis; we can measure growth for sub- and supranational regions. We show, for example, that coastal areas in sub-Saharan Africa are growing slower than the hinterland. (JEL E01, E23, O11, O47, O57)

Luces desde el espacio como proxy de actividad económica

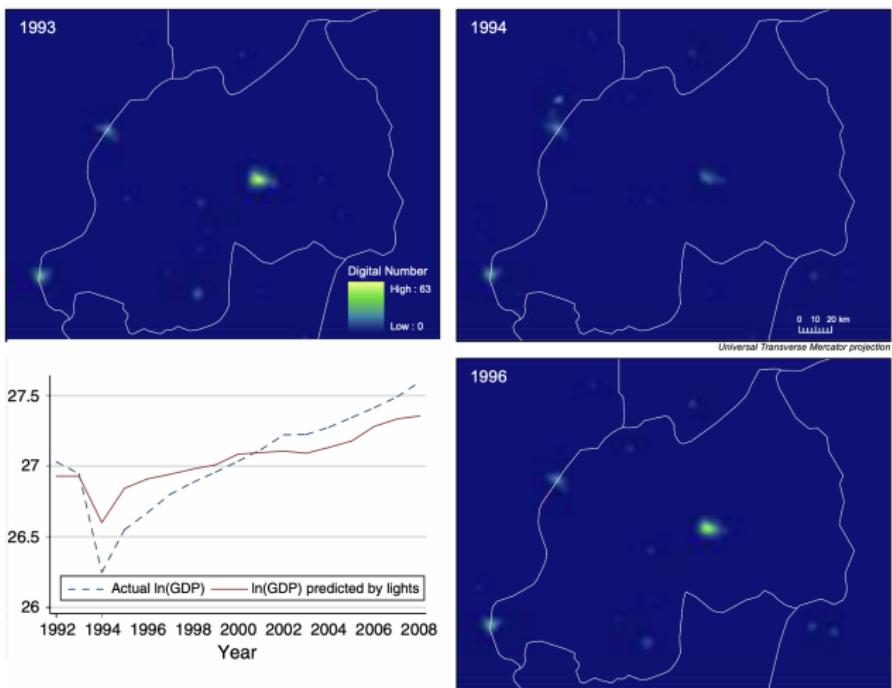


FIGURE 4. GENOCIDE EVENT: RWANDA

Fuente: [Henderson et al. \(2012\)](#)

Literatura adicional sobre PIB y datos satelitales

- Examining the Economic Impact of COVID-19 in India through Daily Electricity Consumption and Nighttime Light Intensity

Autores: Robert C. M. Beyer, Sebastian Franco-Bedoya, Virgilio Galdo

<https://www.sciencedirect.com/science/article/abs/pii/S0305750X2030414>

- Local GDP Estimates Around the World

Autores: Esteban Rossi-Hansberg, Jialing Zhang

<https://bfi.uchicago.edu/working-papers/local-gdp-estimates-around-the-world/>

- How Much Should We Trust the Dictator's GDP Growth Estimates?

Autor: Luis R. Martínez

<https://www.journals.uchicago.edu/doi/10.1086/720458>

1 Motivación

2 Espacio, Datos y Causalidad: Evidencia Reciente

3 Introducción a Datos Espaciales

4 Introducción al Ecosistema de Datos Espaciales en R

¿Qué es un GIS?

- Un **Sistema de Información Geográfica (GIS)** es una tecnología que integra software y hardware para capturar, almacenar, analizar y visualizar información georreferenciada.
- Permite representar fenómenos reales con una dimensión espacial.
- Aplicaciones en urbanismo, salud pública, transporte, medio ambiente, economía urbana, entre otros.

Figure A

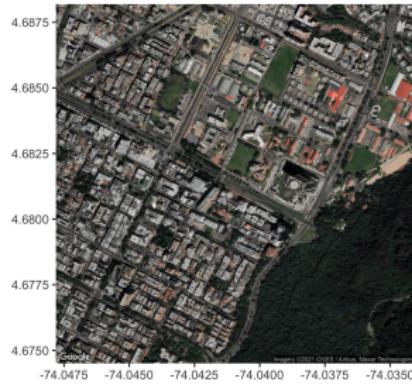


Figure B



Modelos y formatos de datos espaciales

- **Datos vectoriales:**

- ▶ Representan objetos discretos mediante geometrías: puntos, líneas o polígonos.
- ▶ Ejemplos: árboles (punto), calles (línea), barrios (polígono).
- ▶ *Formatos comunes:* .shp, .geojson, .kml, .gpx, .gpkg

- **Datos raster:**

- ▶ Representan variables continuas mediante una grilla de píxeles.
- ▶ Ejemplos: temperatura, altitud, cobertura vegetal.
- ▶ *Formatos comunes:* .tif/.tiff, .asc, .grd, .gri

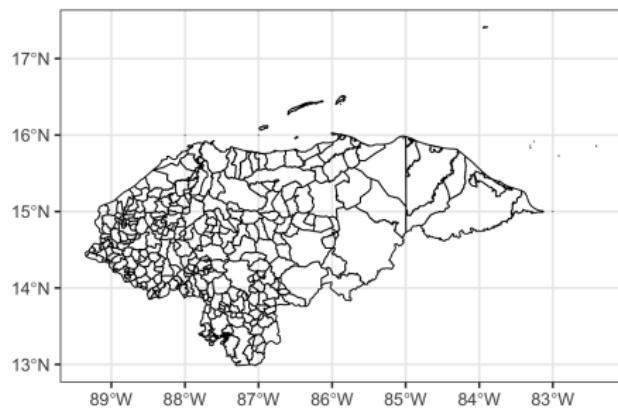
- **Mixtos (vector + raster):**

- ▶ Formatos que pueden almacenar ambos tipos de datos.
- ▶ *Ejemplos:* .sqlite, .gdb, .gpkg

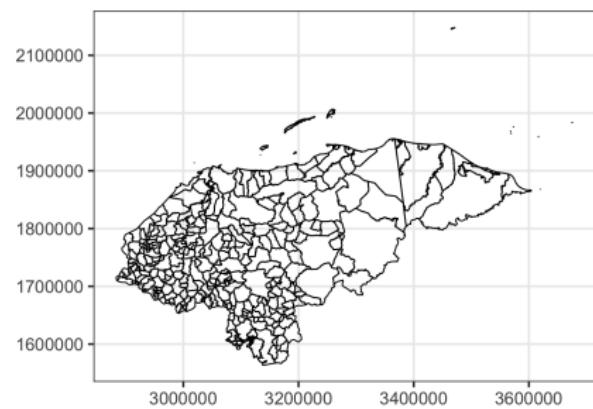
Sistemas de Referencia de Coordenadas (CRS)

- Todo objeto geográfico está asociado a un **CRS** (Coordinate Reference System) que define su ubicación sobre la Tierra.
- Cada CRS tiene un código estandarizado EPSG. Ejemplo: WGS84 = EPSG:4326.
 - ▶ **Geográficos (SCG)**: expresados en grados (latitud, longitud).
 - ▶ **Proyectados (SCP)**: expresados en metros, con proyecciones plana, cónica o cilíndrica.

CRS=WGS84



CRS=NAD83 / UTM zone 12N



Fuentes abiertas de datos espaciales (Colombia y mundo)

Vectoriales:

- **IDECA (Bogotá):** mapas de barrios, manzanas, usos del suelo, equipamientos.
<https://www.ideca.gov.co>
- **IGAC (Colombia):** cartografía básica y catastral a escala 1:25.000 y 1:100.000.
<https://www.igac.gov.co>
- **GeoSIVEL (Colombia):** límites veredales, rurales y rurales dispersos.
<https://geoportal.dane.gov.co>
- **GeoJSON y Shape desde OSM:** a través del paquete osmdata (R) y Overpass API.
Overpass API — osmdata (R)
- **geobr (Brasil, Latinoamérica):** límites municipales y censales desde IBGE.
geobr (R)

Fuentes abiertas de datos espaciales (Colombia y mundo)

Raster:

- **NASA EarthData:** imágenes satelitales MODIS, VIIRS, LANDSAT.
<https://earthdata.nasa.gov>
- **Google Earth Engine (GEE):** acceso gratuito a datos multiespectrales globales.
Compatible con R vía rgee.
- **WorldPop:** mapas de densidad poblacional a alta resolución.
<https://www.worldpop.org>
- **Copernicus / Sentinel:** observación terrestre en tiempo casi real.
<https://scihub.copernicus.eu>

APIs disponibles: osmdata, geobr, rgee, tidygeocoder, elevatr, WFS

1 Motivación

2 Espacio, Datos y Causalidad: Evidencia Reciente

3 Introducción a Datos Espaciales

4 Introducción al Ecosistema de Datos Espaciales en R

Ecosistema de paquetes espaciales en R

- **Paquetes base:**

- ▶ sf: manejo de datos vectoriales con geometría simple.
- ▶ terra: sucesor de raster para datos espaciales tipo raster.
- ▶ sp: estructura antigua para datos espaciales (aún usada).

- **Lectura, escritura y geometría:**

- ▶ rgdal: conexión con la librería GDAL para formatos espaciales.
- ▶ rgeos: operaciones geométricas basadas en GEOS.

- **Visualización:**

- ▶ tmap: mapas temáticos.
- ▶ mapview, leaflet: visualización interactiva.

- **Análisis espacial:**

- ▶ spdep: análisis de dependencia espacial.
- ▶ gstat: interpolación espacial (ej. kriging).
- ▶ osmdata, geobr, tidygeocoder: acceso a datos espaciales externos.

¿Qué es una Simple Feature y por qué usar sf?

- Breve historia:

- ▶ En los 2000s, se utilizaba `sp`, junto con `rgdal` y `rgeos`.
- ▶ Desde 2016, `sf` se consolida como el nuevo estándar:
 - ★ Sintaxis más intuitiva.
 - ★ Mejor rendimiento y visualización.
 - ★ Integra lectura, análisis y escritura de datos espaciales.

- Simple Features (SF) es el estándar definido por la **OGC (Open Geospatial Consortium)**, que permite representar geometrías espaciales (puntos, líneas, polígonos, multipolígonos) de forma interoperable.

- En R, el paquete `sf` implementa este estándar, ofreciendo:

- ▶ Estructura basada en `data.frame` (cada fila = una observación).
- ▶ Columna especial `geometry` que almacena las geometrías.
- ▶ Atributos en columnas regulares; compatible con `tidyverse`.

Más información: <https://r-spatial.github.io/sf/>

¿Cómo luce un objeto sf en R?

- Es un `data.frame` con una columna `geometry`.
- Cada fila representa una entidad geográfica.
- Atributos: columnas tradicionales.
- Geometría: clase `sfc` con CRS asociado.

```
## Simple feature collection with 13767 features and 6 fields  
## Geometry type: POINT  
## Dimension: XY  
## Bounding box: xmin: -75.23494 ymin: 10.31612 xmax: -74.74538 ymax: 11.05187  
## Geodetic CRS: WGS 84  
## First 5 features:
```

	DPTO_CCDGO	MPIO_CCDGO	CPOB_CCDGO	NANG_RTCN	CSIMBOL	geometry
## 1	8	8078	8078001	90	10105	POINT(-74.91679 10.72282)
## 2	8	8078	8078001	90	10105	POINT(-74.9154 10.7253)
## 3	8	8078	8078001	90	10109	POINT(-74.91263 10.73122)
## 4	8	8078	8078001	90	10109	POINT(-74.91318 10.72944)
## 5	8	8078	8078001	90	10109	POINT(-74.91508 10.72809)

simple feature

simple feature geometry list-column

simple feature geometry

¡Pasamos a la parte práctica!

- Hemos finalizado la **introducción teórica al análisis de datos espaciales**.
- Ahora continuaremos con ejercicios prácticos en R usando sf, osmdata, tidygeocoder y otras herramientas.
- Todo el contenido práctico está disponible en el siguiente enlace:

eduard-martinez.github.io

Recuerda tener R y RStudio listos, junto con los paquetes instalados. ¡Vamos al código!

Únete al Semillero CIENFI

- ¿Te interesa la investigación en economía aplicada, datos espaciales, aprendizaje automático o analítica urbana?
- El **Semillero CIENFI** es un espacio para aprender, investigar y trabajar en proyectos reales del Centro de Investigación en Economía y Finanzas.
- ¡Estudiantes de pregrado y posgrado son bienvenidos!

Escanea el siguiente código QR para inscribirte:



Gracias!