

EXPLORANDO COLOMBIA CON GEOPANDAS EN PYTHON

- ♦ **Autor (2025):** Juan Miguel Rodríguez Trujillo, estudiante de economía de noveno semestre
- ♦ **EconomIA:** Laboratorio de inteligencia artificial aplicada a Economía
- ♦ **Universidad Externado de Colombia, Facultad de Economía**



Introducción

La dimensión territorial de los fenómenos económicos es fundamental para comprender las dinámicas de desarrollo, desigualdad y formulación de políticas públicas. A su vez, las herramientas de análisis geoespacial permiten integrar información económica con ubicaciones geográficas, facilitando así la exploración visual de patrones regionales y disparidades socioeconómicas.

En este contexto, Python ofrece poderosas bibliotecas para el análisis de datos espaciales. Una de las más destacadas es GeoPandas, que extiende la funcionalidad de pandas para manipular objetos geográficos de forma intuitiva, permitiendo cargar, unir y visualizar datos con componentes espaciales. Su facilidad de integración con matplotlib hace de GeoPandas una excelente herramienta para la elaboración de mapas temáticos y el análisis regional.

Este artículo presenta una guía práctica para explorar las desigualdades regionales en Colombia utilizando GeoPandas. A través de ejemplos aplicados, se visualizan cinco indicadores clave a nivel departamental:

- Índice de Pobreza Multidimensional (IPM)
- Población con educación universitaria
- PIB per cápita
- Tasa de desempleo
- Índice de Competitividad Departamental (IDC)

Los datos provienen de fuentes oficiales como el Departamento Administrativo Nacional de Estadística (DANE) y el Consejo Privado de Competitividad (CPC), y se vinculan con un archivo geoespacial en formato shapefile que contiene los límites departamentales de Colombia.

Como complemento didáctico, también se presenta un ejercicio de visualización global utilizando GeoPandas, donde se presenta el Índice de Gini por país en un mapamundi. Este ejercicio tiene como propósito ilustrar el potencial de GeoPandas más allá del ámbito nacional, abriendo la posibilidad de realizar comparaciones internacionales y visualizar desigualdades globales.

A lo largo del documento, cada mapa generado será precedido por una breve explicación metodológica y acompañado por una interpretación económica de los resultados. De esta manera, el lector no solo aprenderá el uso técnico de GeoPandas, sino que también desarrollará una comprensión analítica de los patrones espaciales presentes en los datos económicos regionales de Colombia.

Importar librerías

Importamos las librerías necesarias para realizar el análisis: `pandas`, `geopandas` y `matplotlib`

```
In [7]: import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt

from matplotlib.lines import Line2D
from matplotlib.patches import Patch
```

Cargar archivos .xlsx con información de las variables socioeconómicas para Colombia

Cargamos archivos .xlsx con la librería `pandas`

1. Índice de pobreza multidimensional (IPM)

```
In [11]: df_ipm = pd.read_excel(ruta_archivo_IPM)
```

2. Población con estudios de pregrado o superiores

```
In [13]: df_universitarios = pd.read_excel(ruta_archivo_universitarios)
```

3. PIB per cápita

```
In [15]: df_PIB = pd.read_excel(ruta_archivo_PIB_per_capita)
```

4. Tasa de desempleo

```
In [17]: df_desempleo = pd.read_excel(ruta_archivo_desempleo)
```

5. Índice de competitividad (IDC)

```
In [19]: df_IDC = pd.read_excel(ruta_archivo_IDC)
```

Estandarizar nombres de los elementos de la columna "DEPARTAMENTOS"

Es fundamental homogenizar los nombres de los elementos de la columna identificadora "DEPARTAMENTOS" para posteriormente realizar el merge. En este caso los estandarizamos usando mayúsculas y eliminando espacios extra innecesarios en la escritura. Este proceso se realiza para todos los datasets.

```
In [22]: df_ipm['DEPARTAMENTO'] = df_ipm['DEPARTAMENTO'].str.upper().str.strip()
df_universitarios['DEPARTAMENTO'] = df_universitarios['DEPARTAMENTO'].str.upper().s
df_PIB['DEPARTAMENTO'] = df_PIB['DEPARTAMENTO'].str.upper().str.strip()
df_desempleo['DEPARTAMENTO'] = df_desempleo['DEPARTAMENTO'].str.upper().str.strip()
df_IDC['DEPARTAMENTO'] = df_IDC['DEPARTAMENTO'].str.upper().str.strip()
```

Cargar archivo .shapefile con información geográfica de los departamentos de Colombia

Un archivo Shapefile es un formato estándar utilizado en los Sistemas de Información Geográfica (SIG) para almacenar datos geospaciales vectoriales. Un Shapefile puede representar puntos, líneas o polígonos para mapear ubicaciones geográficas como ciudades, ríos o límites políticos.

Aunque se le llama "archivo Shapefile", en realidad está compuesto por varios archivos que trabajan juntos, siendo los principales:

- shp — Contiene las geometrías (formas) de los objetos geográficos.
- shx — Es un índice que facilita el acceso a los datos.
- dbf — Almacena los atributos asociados (datos tabulares) de cada objeto geográfico.

Para cargar el archivo .shapefile hacemos uso de **geopandas**. Esto permite interpretar los elementos de las variables del archivo .shapefile como coordenadas y límites geográficos.

El DANE proporciona para Colombia el archivo .shapefile de libre acceso

```
In [27]: shapefile_colombia_ruta = r"C:\Users\juanr\OneDrive\Escritorio\Proyectos Individual
colombia_mapa = gpd.read_file(shapefile_colombia_ruta)
```

Gráfico del mapa base para Colombia

```
In [29]: colombia_mapa.plot(figsize=(10, 8), edgecolor='black', facecolor='none')
plt.title('Mapa base de Colombia', fontsize=14)
plt.axis('off') # Opcional para eliminar los ejes
plt.show()
```

Mapa base de Colombia



Estandarizar elementos de variable identificadora para el .shapefile

En esta parte estandarizamos los nombres de los departamentos. Añadimos mayúscula sostenida y eliminamos espacios innecesarios en su escritura con las funciones `.str.upper()` y `.str.strip()`, respectivamente.

También modificamos los nombres de algunos departamentos para que coincidan con los `df_xxxx`. Este se logra con la función `.replace`.

Finalmente, renombramos la columna `'dpto_cnomb'` y la nombramos `'DEPARTAMENTO'`. Esto con la función `.rename`.

```
In [32]: colombia_mapa["dpto_cnabr"] = colombia_mapa["dpto_cnabr"].str.upper().str.strip()

colombia_mapa['dpto_cnabr'] = colombia_mapa['dpto_cnabr'].replace({
    'QUINDIO': 'QUINDÍO',
    'BOGOTÁ, D.C.': 'BOGOTÁ D.C.',
    'ARCHIPIÉLAGO DE SAN ANDRÉS, PROVIDENCIA Y SANTA CATALINA': 'SAN ANDRÉS'
})

colombia_mapa.rename(columns={'dpto_cnabr': 'DEPARTAMENTO'}, inplace=True)
```

Índice de Pobreza Multidimensional (IPM)

Exploraremos primero el Índice de Pobreza Multidimensional (IPM), una medida integral que combina indicadores de salud, educación y condiciones de vida. Este índice es fundamental para identificar territorios con múltiples privaciones simultáneas. El objetivo de esta visualización es mostrar cómo la pobreza estructural varía entre los departamentos del país.

Merge

En este caso usamos datos de la **Encuesta Nacional de Calidad de Vida (ECV) - 2023** realizada por el **DANE**. Luego, realizamos el **merge de datasets**: `colombia_mapa`, que contiene información geográfica de Colombia, y `df_ipm`, que contiene información sobre el IPM, generando el nuevo dataset `colombia_ipm`.

```
In [37]: colombia_ipm = colombia_mapa.merge(df_ipm, on='DEPARTAMENTO', how='left')
```

Gráfico

Realizamos el gráfico con información de la columna `IPM (Porcentaje)`, del dataset `colombia_ipm`.

```
In [40]: fig, ax = plt.subplots(1, 1, figsize=(12, 8))

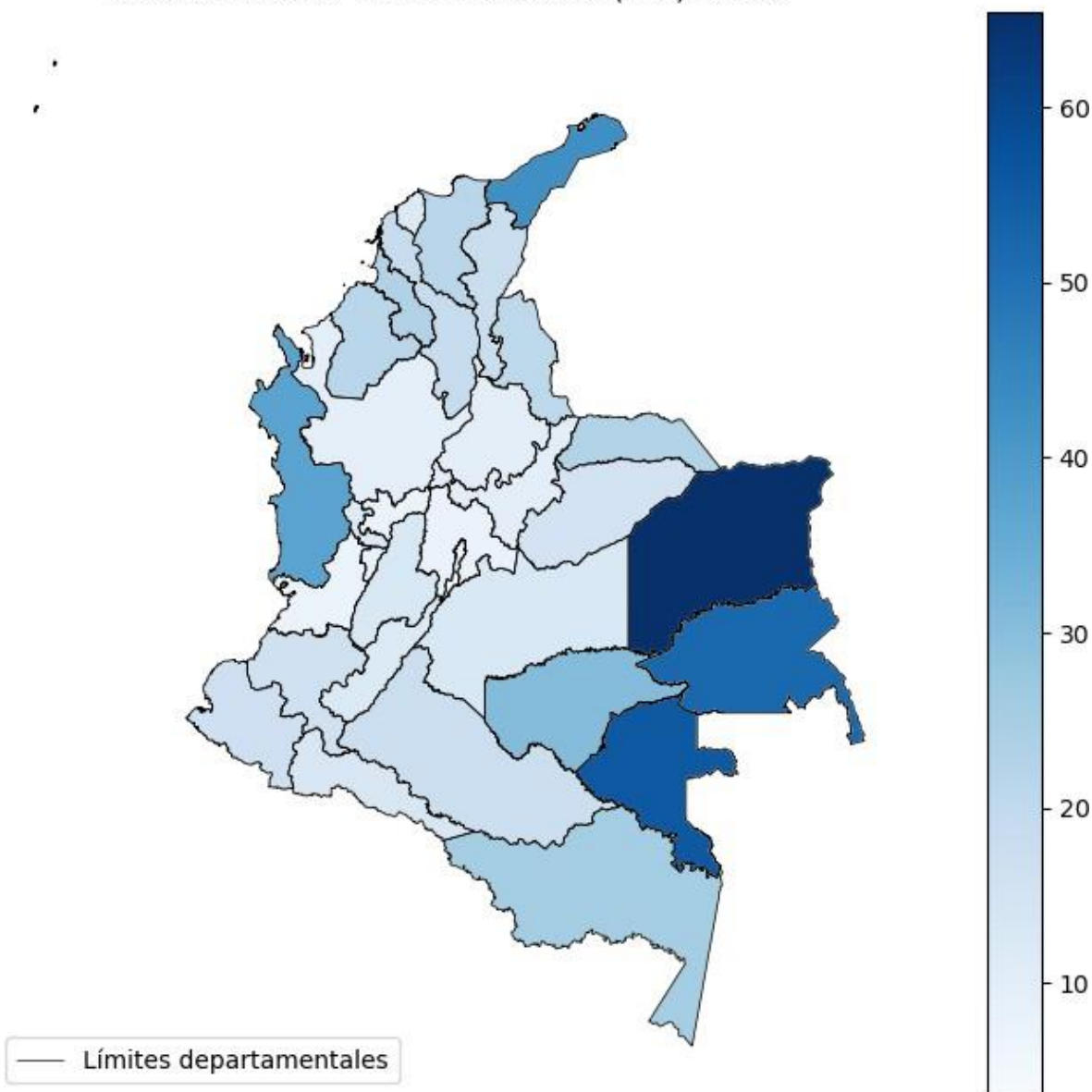
colombia_ipm.plot(column='IPM (Porcentaje)', cmap='Blues', linewidth=0.8, ax=ax, ed
    "color": "lightgrey",
    "edgecolor": "red",
    "hatch": "///",
    "label": "Sin datos"
})

colombia_ipm.boundary.plot(ax=ax, color='black', linewidth=0.5, label='Límites depa

ax.legend(loc='lower left')
ax.set_title('Tasa de Pobreza Multidimensional (IPM) - 2023')
ax.axis('off')
```

```
plt.show()
```

Tasa de Pobreza Multidimensional (IPM) - 2023



Este mapa permite identificar claramente la persistencia de desigualdades estructurales en Colombia. Se observa que los departamentos de la Amazonía y la Orinoquía —como Vaupés, Guainía, Guaviare y Vichada— presentan las tasas más altas de pobreza multidimensional, superando en algunos casos el 50%. Estas regiones históricamente han enfrentado problemas de aislamiento geográfico, debilidad institucional y baja cobertura de servicios básicos, lo que contribuye a la acumulación de privaciones en dimensiones como educación, salud, acceso a agua potable y condiciones de vivienda.

En contraste, los departamentos del centro y occidente del país, especialmente aquellos con economías más diversificadas y mayor urbanización como Bogotá D.C., Cundinamarca, Antioquia y el Valle del Cauca, muestran niveles considerablemente más bajos de IPM, reflejando mejores condiciones de vida y mayor cobertura de servicios públicos.

Población con estudios universitarios de pregrado

La proporción de población con educación universitaria refleja las oportunidades de formación superior y el desarrollo del capital humano en cada región. Mapear esta variable permite identificar las diferencias en acceso a la educación de calidad, así como su posible correlación con otros indicadores como competitividad departamental, desempleo o ingresos.

Merge

En este caso usamos datos del **Censo Nacional de Población y Vivienda (CNPV) - 2018** realizado por el **DANE**. Luego, realizamos el **merge de datasets**: `colombia_mapa`, que contiene información geográfica de Colombia, y `df_universitarios`, que contiene información sobre los estudios de la población colombiana, generando el nuevo dataset `colombia_universitarios`.

```
In [45]: colombia_universitarios = colombia_mapa.merge(df_universitarios, on='DEPARTAMENTO',
```

Gráfico

Realizamos el gráfico con información de la columna `Universitarios`, del dataset `colombia_universitarios`.

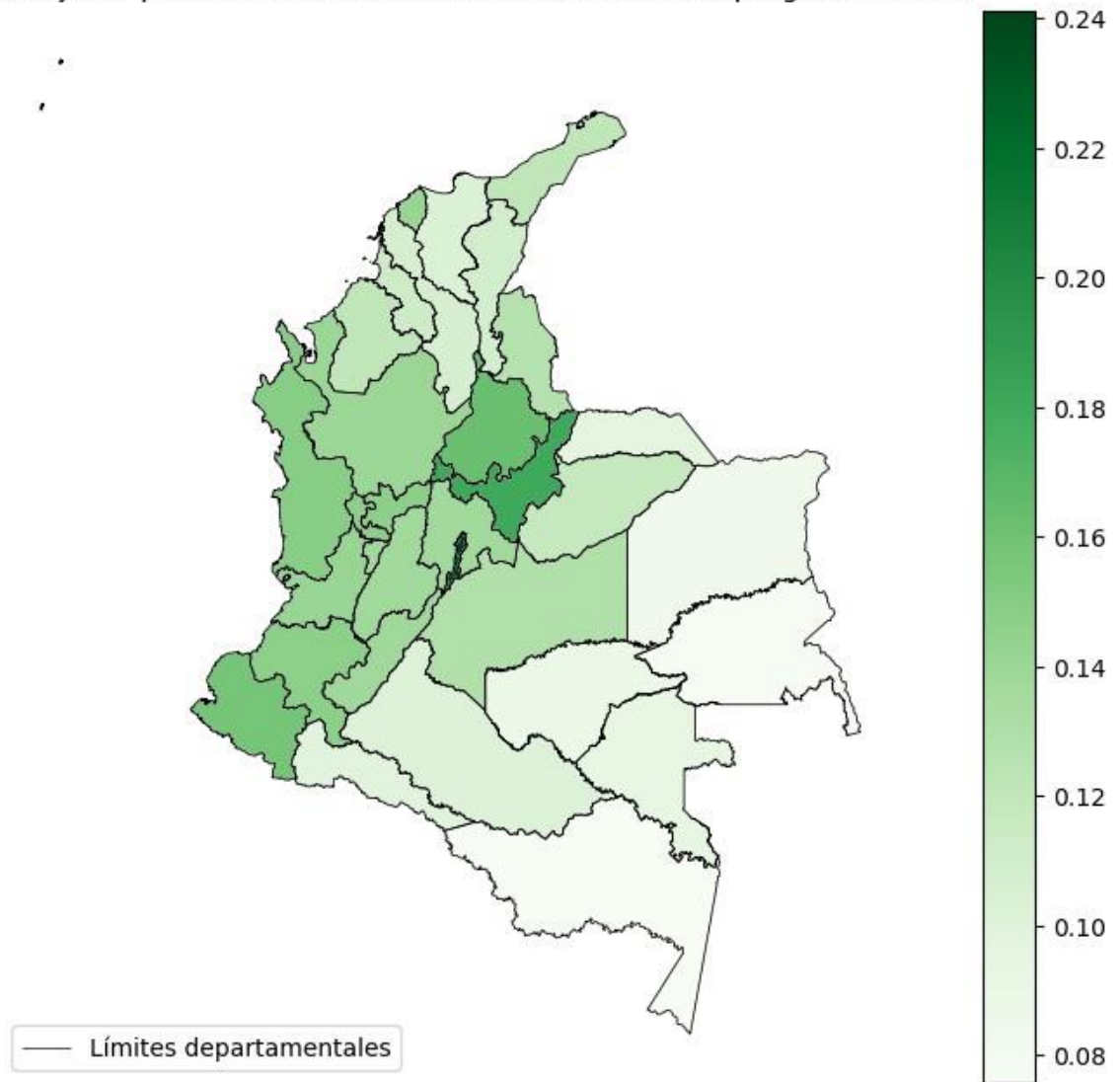
```
In [48]: fig, ax = plt.subplots(1, 1, figsize=(12, 8))

colombia_universitarios.plot(column='Universitarios', cmap='Greens', linewidth=0.8,
    "color": "lightgrey",
    "edgecolor": "red",
    "hatch": "///",
    "label": "Sin datos"
})

colombia_universitarios.boundary.plot(ax=ax, color='black', linewidth=0.5, label='L
ax.legend(loc='lower left')
ax.set_title('Porcentaje de población con estudios universitarios de pregrado - 201
ax.axis('off')

plt.show()
```

Porcentaje de población con estudios universitarios de pregrado - 2018



Este mapa refleja marcadas diferencias en el nivel educativo superior alcanzado por la población en los distintos departamentos de Colombia. Destaca Bogotá D.C. como el territorio con la mayor proporción de personas con estudios universitarios, consolidándose como el principal centro educativo del país. Esta concentración obedece a su histórica oferta de instituciones de educación superior, mejores condiciones laborales para profesionales y una alta demanda de formación calificada.

En contraste, amplias zonas del sur y oriente —como Vaupés, Guainía o Amazonas— muestran niveles significativamente más bajos, evidenciando barreras persistentes de acceso territorial, económico y cultural a la educación superior.

PIB per cápita

El Producto Interno Bruto per cápita departamental es una medida clásica del ingreso promedio y desarrollo económico relativo. Representarlo en el territorio revela la concentración de la riqueza y permite analizar desequilibrios productivos entre regiones, clave para diseñar políticas de convergencia territorial. En Colombia, el DANE calcula el PIB per cápita:

$$\text{PIB per cápita} = \frac{\text{PIB}}{\text{Población total}}$$

Merge

En este caso usamos datos del **PIB total departamental (2023)**, calculado por el **DANE**. Luego, realizamos el **merge de datasets**: `colombia_mapa`, que contiene información geográfica de Colombia, y `df_PIB`, que contiene información sobre el valor del PIB departamental, generando el nuevo dataset `colombia_PIB`.

```
In [53]: colombia_PIB = colombia_mapa.merge(df_PIB, on='DEPARTAMENTO', how='left')
```

```
In [56]: fig, ax = plt.subplots(1, 1, figsize=(12, 8))

colombia_PIB.plot(column='PIB per cápita', cmap='Reds', linewidth=0.8, ax=ax, edgec
    "color": "lightgrey",
    "edgecolor": "red",
    "hatch": "///",
    "label": "Sin datos"
})

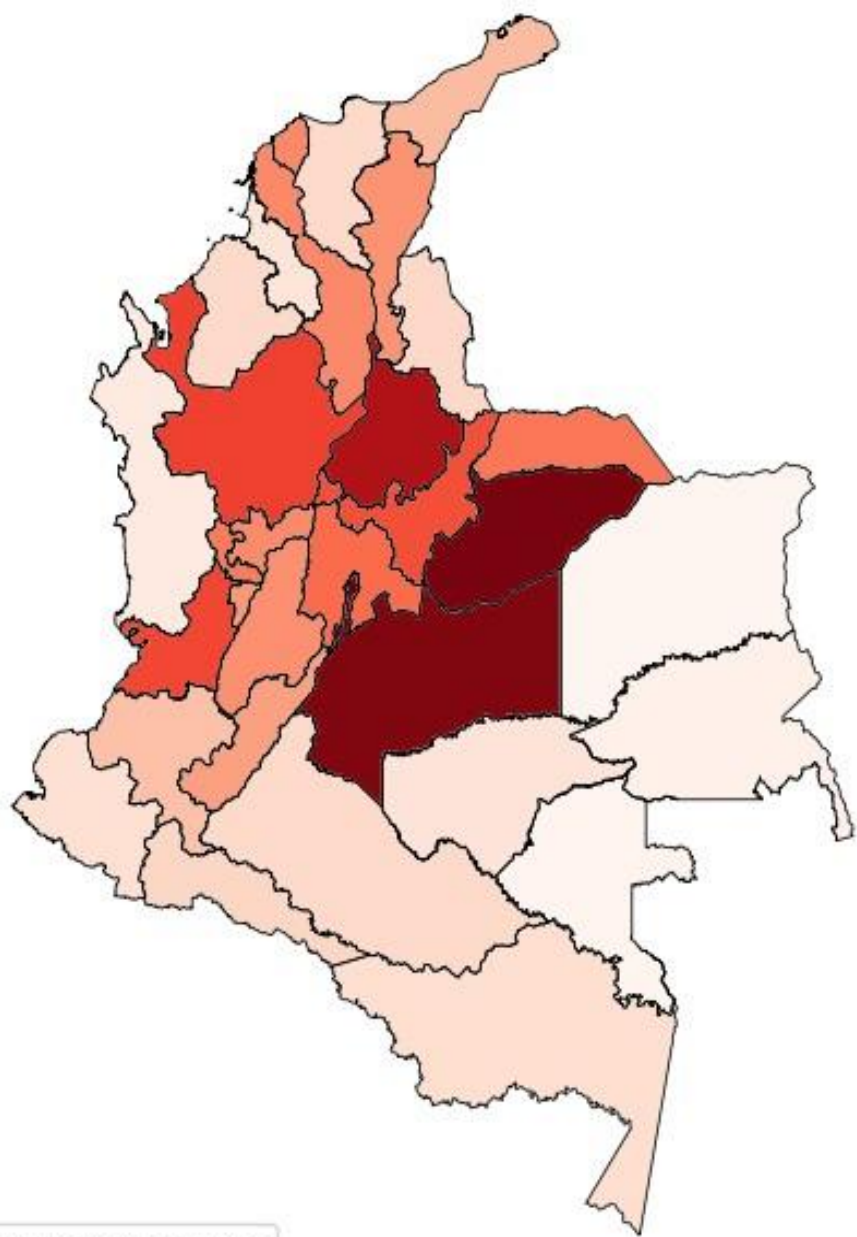
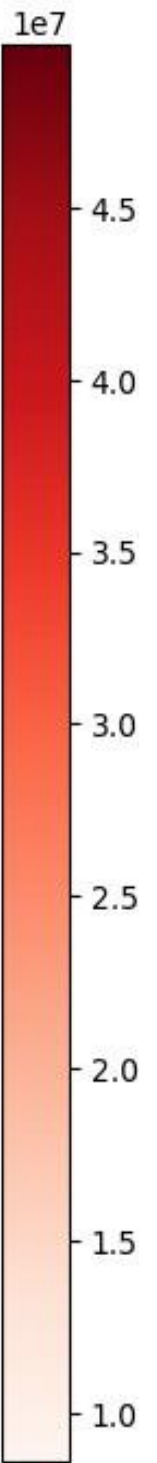
colombia_PIB.boundary.plot(ax=ax, color='black', linewidth=0.5, label='Límites depa
ax.legend(loc='lower left')
ax.set_title('PIB per cápita - 2023')
ax.axis('off')

plt.show()
```

Gráfico

Realizamos el gráfico con información de la columna `PIB per cápita`, del dataset `colombia_PIB`.

PIB per cápita - 2023



— Límites departamentales

Este mapa evidencia una alta concentración del PIB per cápita en el centro y oriente del país, particularmente en Meta y Casanare. Esta situación se explica por la fuerte presencia de actividades extractivas, especialmente la explotación petrolera, que eleva el ingreso promedio en relación con el tamaño poblacional.

En contraste, departamentos como Chocó, Nariño y los de la Amazonía presentan niveles significativamente más bajos, reflejando estructuras económicas menos productivas y menor inserción en cadenas de valor dinámicas.

Tasa de Desempleo

La tasa de desempleo departamental permite observar la situación del mercado laboral en cada región. Visualizarla geográficamente ofrece una perspectiva más clara de las brechas laborales, posibles rezagos estructurales y retos de inclusión productiva. Este mapa ayudará a identificar territorios donde la generación de empleo es una prioridad crítica. En Colombia es calculada por el DANE con la siguiente fórmula:

$$TD = \frac{DS}{FT} \times 100$$

Donde :

TD: Tasa de desempleo	(8)
DS: Población desempleada	(9)
FT: Fuerza de trabajo	(10)

Merge

En este caso usamos datos de la **Gran Encuesta Integrada de Hogares (GEIH) - 2023**, realizada por el **DANE**. Luego, realizamos el **merge de datasets**: `colombia_mapa`, que contiene información geográfica de Colombia, y `df_desempleo`, que contiene información sobre la tasa de desempleo, generando el nuevo dataset `colombia_desempleo`.

```
In [61]: colombia_desempleo = colombia_mapa.merge(df_desempleo, on='DEPARTAMENTO', how='left')
```

Gráfico

Realizamos el gráfico con información de la columna `Desempleo`, del dataset `colombia_desempleo`.

```
In [64]: fig, ax = plt.subplots(1, 1, figsize=(12, 8))

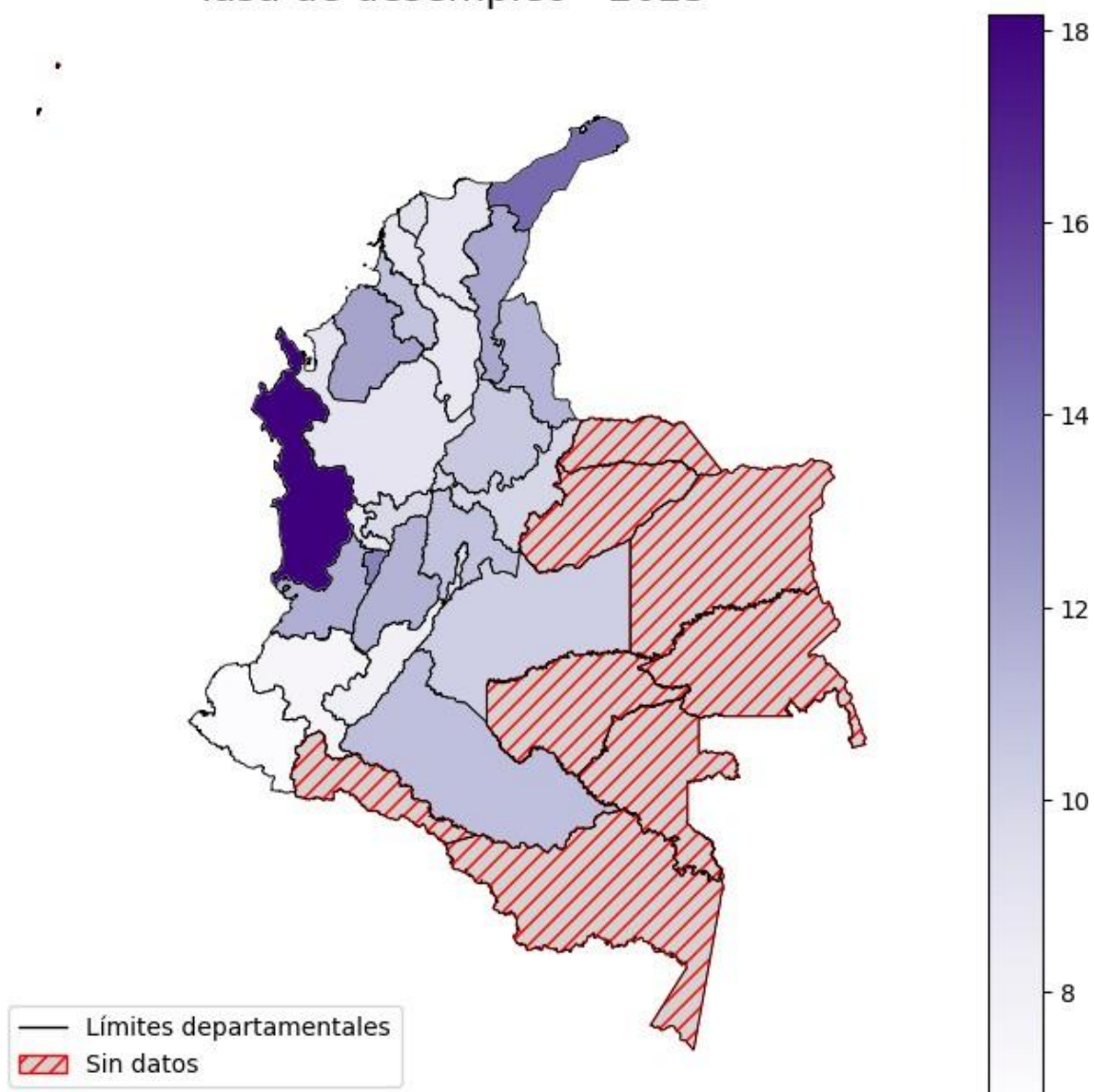
colombia_desempleo.plot(column='Desempleo', cmap='Purples', linewidth=0.8, ax=ax, e
    "color": "lightgrey",
    "edgecolor": "red",
    "hatch": "///",
    "label": "Sin datos"
})

colombia_desempleo.boundary.plot(ax=ax, color='black', linewidth=0.5)

legend_elements = [
    Line2D([0], [0], color='black', lw=1, label='Límites departamentales'),
    Patch(facecolor='lightgrey', edgecolor='red', hatch='///', label='Sin datos')
]

ax.legend(handles=legend_elements, loc='lower left')
ax.set_title('Tasa de desempleo - 2023', fontsize=16)
ax.axis('off')
plt.show()
```

Tasa de desempleo - 2023



Este mapa revela disparidades marcadas en el acceso al empleo en Colombia. Destacan con preocupación Chocó y La Guajira, departamentos con las tasas de desempleo más elevadas, lo cual refleja una combinación de baja diversificación productiva y déficits históricos en infraestructura y capital humano.

En contraste, departamentos del suroccidente como Huila, Cauca y Nariño exhiben un desempeño relativamente favorable, con tasas de desempleo por debajo del promedio nacional. Aunque enfrentan retos en otras dimensiones socioeconómicas, estas regiones muestran avances en inclusión laboral. También destacan positivamente Bogotá D.C. y Antioquia por sus bajas tasas de desempleo.

Índice de Competitividad Departamental (IDC)

El IDC resume factores institucionales, económicos y sociales que determinan la capacidad de un departamento para generar bienestar y crecimiento sostenible. Este indicador compila dimensiones como infraestructura, educación, entorno empresarial y sofisticación productiva. Su visualización permite identificar regiones con mayores ventajas estructurales y aquellas con rezagos que requieren intervención estratégica.

Merge

En este caso usamos datos de los resultados del **Informe: Índice de Competitividad Departamental (2024)**, realizado por el **CPC**. Luego, realizamos el **merge de datasets**: `colombia_mapa`, que contiene información geográfica de Colombia, y `df_IDC`, que contiene información sobre la tasa de desempleo, generando el nuevo dataset `colombia_IDC`.

```
In [69]: colombia_IDC = colombia_mapa.merge(df_IDC, on='DEPARTAMENTO', how='left')
```

Gráfico

Realizamos el gráfico con información de la columna `IDC`, del dataset `colombia_IDC`.

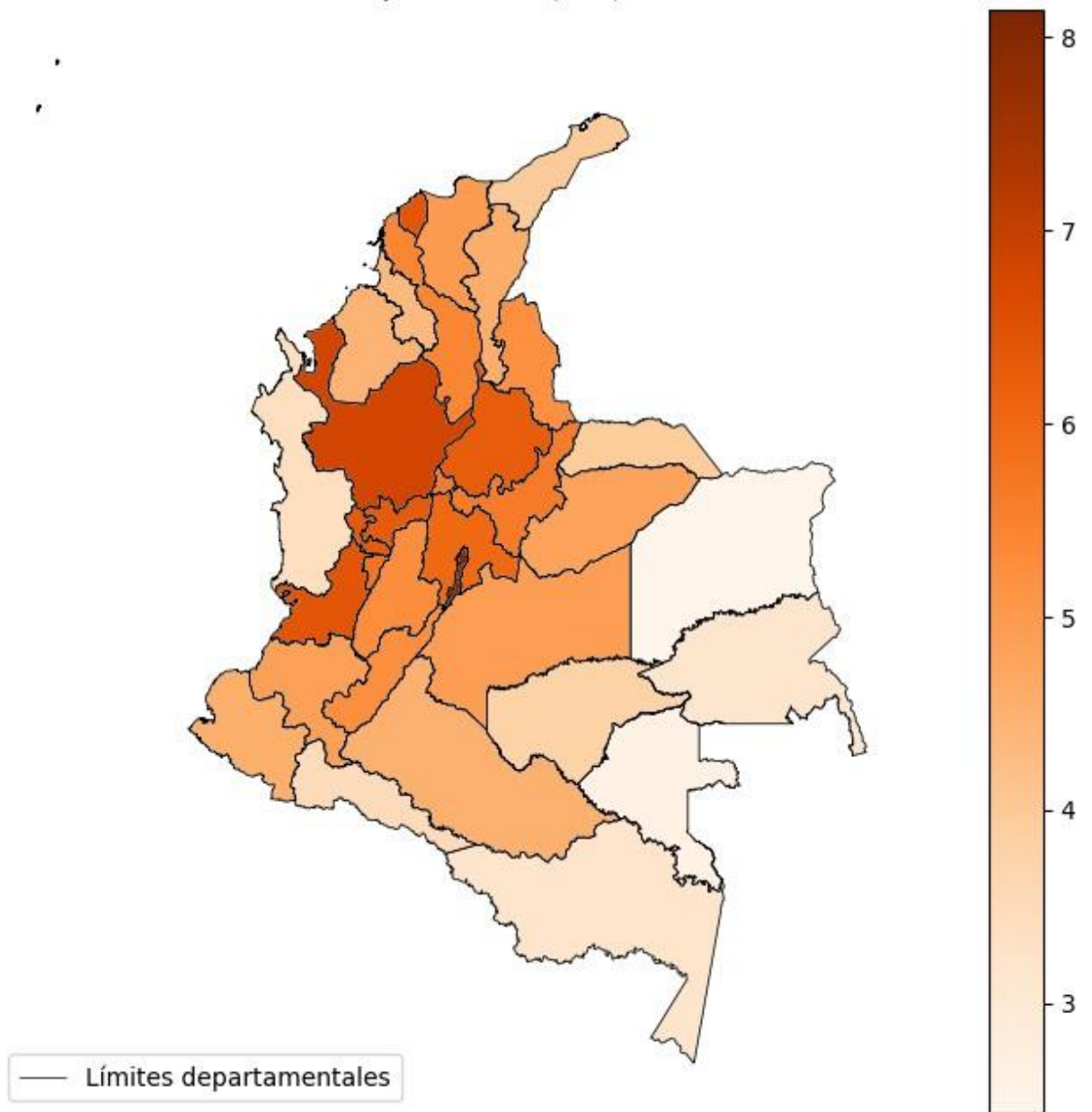
```
In [72]: fig, ax = plt.subplots(1, 1, figsize=(12, 8))

colombia_IDC.plot(column='IDC', cmap='Oranges', linewidth=0.8, ax=ax, edgecolor='0.
    "color": "lightgrey",
    "edgecolor": "red",
    "hatch": "///",
    "label": "Sin datos"
})

colombia_ipm.boundary.plot(ax=ax, color='black', linewidth=0.5, label='Límites de pa
ax.legend(loc='lower left')
ax.set_title('Índice de competitividad (IDC) - 2024')
ax.axis('off')

plt.show()
```

Índice de competitividad (IDC) - 2024



El mapa evidencia un claro patrón de concentración de la competitividad en el centro del país, destacando Bogotá D.C., Antioquia, Valle del Cauca, Santander y Cundinamarca, que obtienen los mejores puntajes en el IDC. Este índice refleja capacidades institucionales, infraestructura, acceso a educación, entorno empresarial, innovación y sofisticación del aparato productivo.

La competitividad más baja se observa en regiones periféricas como el suroriente del país y el Chocó, lo que denota desigualdades estructurales en el acceso a conectividad, calidad institucional y oportunidades económicas

GeoPandas para el Mapamundi

Como ejercicio complementario, se presenta un mapamundi del Índice de Gini por país. Esta medida de desigualdad permite comparar globalmente las diferencias en distribución del ingreso. A través de GeoPandas, es posible generar mapas a escala mundial que contextualicen el caso colombiano dentro de un panorama internacional.

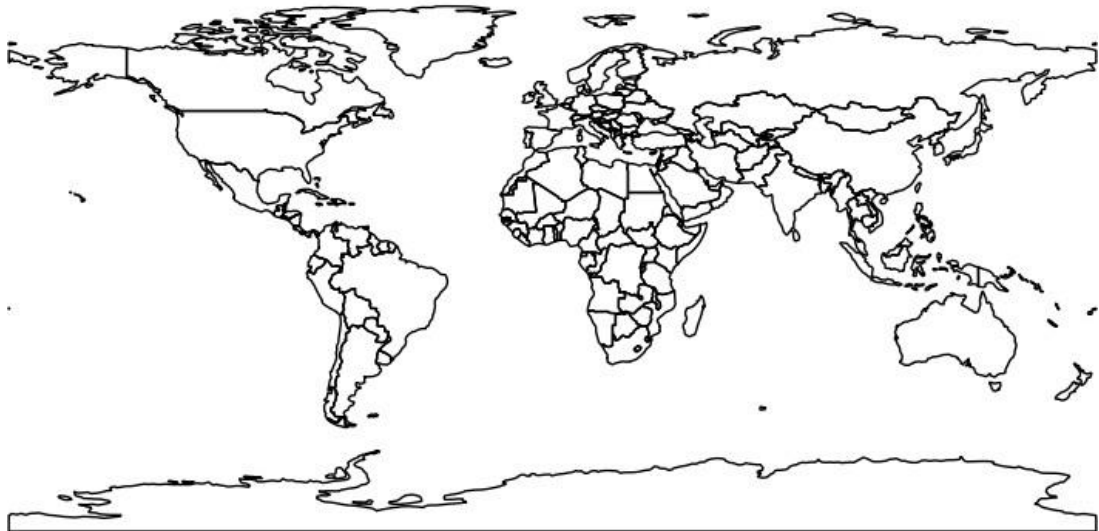
Shapefile de Natural Earth

```
In [76]: # Ruta al archivo .shapefile descargado de Natural Earth
shapefile_path_world = r"ruta_archivo_shapefile"
# Cargar el Shapefile con GeoPandas
world = gpd.read_file(shapefile_path_world)

# Estandarizar los nombres de los países en el Shapefile: mayúsculas y sin espacios
world["ADMIN"] = world["ADMIN"].str.upper().str.strip()

# Generar Mapa base del mundo
world.plot(figsize=(10, 8), edgecolor='black', facecolor='none')
plt.title('Mapa base del mundo', fontsize=14)
plt.axis('off')
plt.show()
```

Mapa base del mundo



Dataset del World Bank

```
In [78]: # Cargar datos del Banco Mundial sobre el índice GINI
ruta_archivo_GINI = r"ruta_archivo_GINI"
df_GINI = pd.read_excel(ruta_archivo_GINI)

# Estandarizar los nombres de los países en el dataset del GINI (mayúsculas y sin e
df_GINI['ADMIN'] = df_GINI['ADMIN'].str.upper().str.strip()

# Realizar el merge
world_GINI = world.merge(df_GINI, on='ADMIN', how='left')
```

Definición del Índice de GINI

El Índice de Gini es una medida estadística que cuantifica la desigualdad en la distribución del ingreso o la riqueza dentro de una población. Su valor oscila entre 0 y 1, donde 0 representa igualdad perfecta (todos tienen los mismos ingresos) y 1 indica desigualdad máxima (una sola persona concentra todos los ingresos). Es ampliamente utilizado para analizar disparidades económicas y evaluar el nivel de equidad en una sociedad.

Formalmente se puede representar como:

$$IG = \frac{\sum_{i=1}^{k-1} (p_i - q_i)}{\sum_{i=1}^{k-1} p_i}$$

Donde:

IG : Índice de Gini

p_i : Proporción de individuos con ingresos menores a una cantidad determinada

q_i : Porcentaje de ingresos acumulado sobre el total de los ingresos de la población

$i = 1, 2, \dots, k$: Conjunto de individuos o grupos poblacionales analizados

k : Número total de individuos o grupos considerados

Gráfico

```

In [82]: # Realizar gráfico a partir del DataSet creado "world_GINI"

fig, ax = plt.subplots(1, 1, figsize=(12, 8))

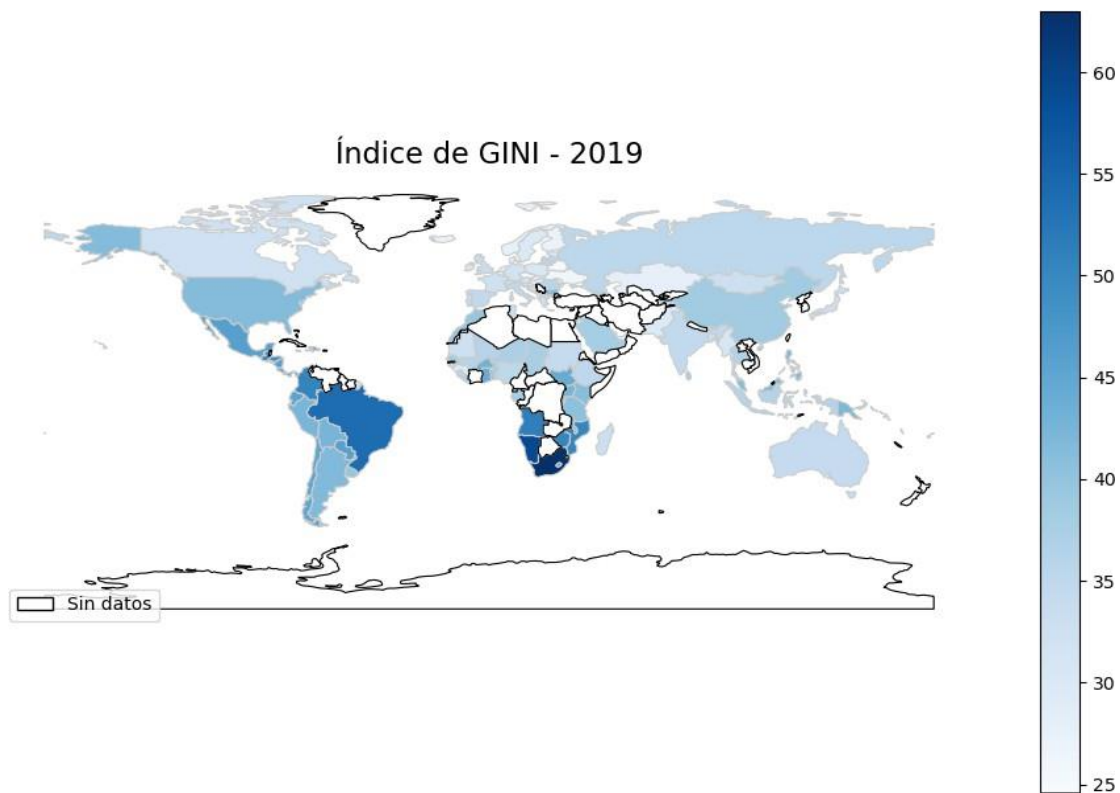
world_GINI.plot(column='GINI', cmap='Blues', linewidth=0.8, ax=ax, edgecolor='0.8',
                "color": "white",
                "edgecolor": "black"
                })

legend_elements = [
    Patch(facecolor='white', edgecolor='black', label='Sin datos')]

ax.legend(handles=legend_elements, loc='lower left')
ax.set_title('Índice de GINI - 2019', fontsize=16)
ax.axis('off')

plt.show()

```



El mapa global del Índice de Gini muestra con claridad la elevada desigualdad de ingresos que caracteriza a América Latina, consolidándola como una de las regiones más inequitativas del mundo. Colombia y Brasil, en particular, destacan por tener coeficientes superiores a 50, reflejo de persistentes brechas estructurales en acceso a educación, empleo formal y activos productivos. La situación es aún más crítica en países del sur de África como Sudáfrica y Namibia, que registran los niveles más altos de desigualdad a nivel mundial.

En contraste, los países nórdicos (como Noruega, Suecia y Finlandia) se posicionan como referentes de equidad distributiva, con niveles de Gini alrededor de 30.

Conclusión

Esta guía ha demostrado el valor del análisis geoespacial como herramienta fundamental para la exploración de desigualdades económicas y sociales en Colombia. Mediante el uso de la biblioteca GeoPandas en Python, fue posible integrar información estadística con datos geográficos para construir visualizaciones temáticas que revelan patrones regionales en dimensiones clave del desarrollo. A través de los mapas del IPM, desempleo, educación superior, PIB per cápita e Índice de Competitividad Departamental, se evidenció la persistencia de brechas territoriales estructurales que inciden en el bienestar y las oportunidades de la población. Además, el ejercicio complementario a escala global con el Índice de Gini permitió contextualizar la situación colombiana dentro de un panorama internacional de desigualdad.

Más allá de los aspectos técnicos, esta guía busca promover una lectura crítica del territorio como unidad de análisis económico. Visualizar datos sobre un mapa no solo facilita su interpretación, sino que permite pensar la política pública desde una perspectiva diferencial, focalizada y basada en evidencia.

La combinación entre análisis de datos y economía territorial, habilitada por herramientas como GeoPandas, ofrece un camino prometedor para investigadores, estudiantes y analistas interesados en comprender y transformar realidades locales desde un enfoque visual, riguroso y reproducible.