

Análisis espacial avanzado

I de Moran

I de Moran Global

Resume en un coeficiente el grado de asociación entre un nivel dado de la variable considerada en un área geográfica respecto del promedio ponderado de la misma variable en áreas vecinas

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right)}$$

Moran, P. A. P. (1950) Notes on continuous stochastic phenomena. Biometrika, 37, 17–23.
Cliff, A. D. and Ord, J. K. (1973) Spatial Autocorrelation. London: Pion.

Autocorrelación espacial

“Todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes.”

Ley de Tobler

TOBLER, W. R (1970). *A computer model simulation of urban growth in the Detroit region*. Economic Geography 46(2).

Esto nos permite observar el fenómeno de la dependencia espacial o autocorrelación espacial

Haining, R. (2003) *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

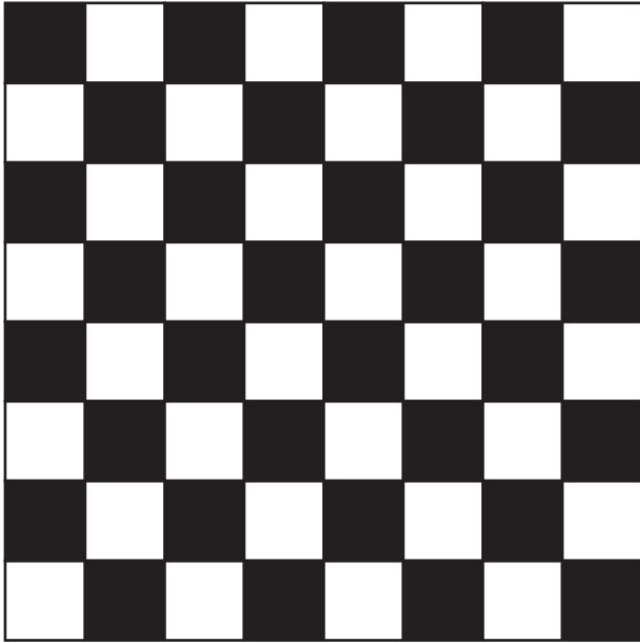
Autocorrelación espacial

El grado en el cual una variable está espacialmente correlacionada con ella misma. Puede ser que todos los valores sean similares (autocorrelación positiva) o totalmente independientes (todos los valores son disímiles entre sí)

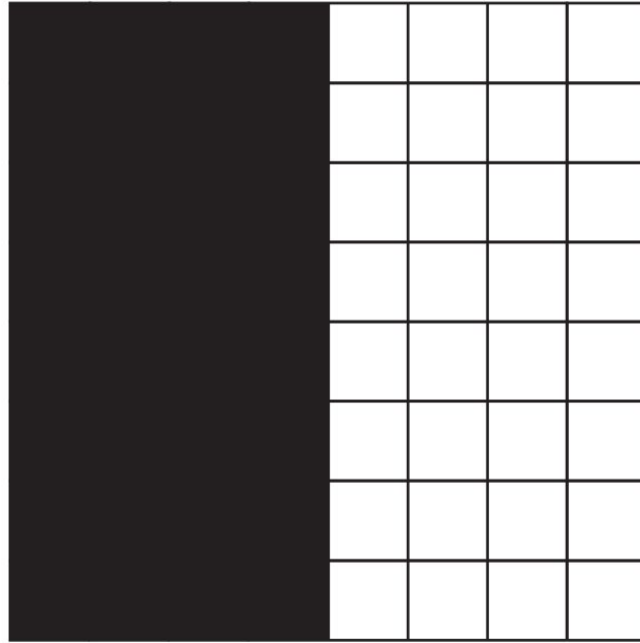
Haining, R. (2003) Spatial Data Analysis: Theory and Practice. Cambridge: Cambridge University Press.

También es utilizado (en su variante Local) para detectar clusters espaciales. En sus 4 variantes. Más adelante veremos eso...

Resultados del índice de Moran

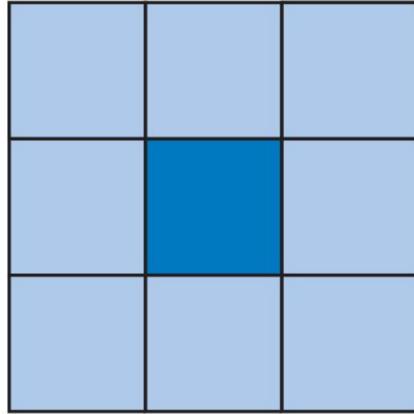


Negative spatial autocorrelation
 $I = -1.000$

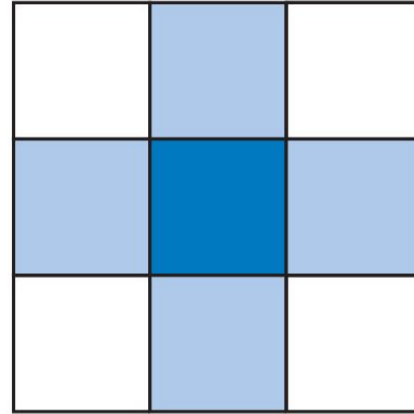


Positive spatial autocorrelation
 $I = 0.865$

Adyacencias tipo Torre o tipo Reina



Queen's case
contiguity



Rook's case
contiguity



A

45	44	44	43	42	44	38	42
43	42	34	35	41	42	42	39
38	32	34	35	38	42	41	39
39	41	37	38	39	42	41	39
32	41	35	31	39	38	37	41
29	29	28	33	41	44	43	42
34	36	37	39	41	37	39	41
35	37	36	36	39	41	36	42

B

Moran's I : Rook's case: 0.505
Queen's case: 0.400

Ecuación de I de Moran Global

7	8	11
11	9	10
11	12	9

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right)}$$

$$I = \frac{9 \times 3.975}{21.556 \times 40} = \frac{35.778}{862.222} = 0.041$$

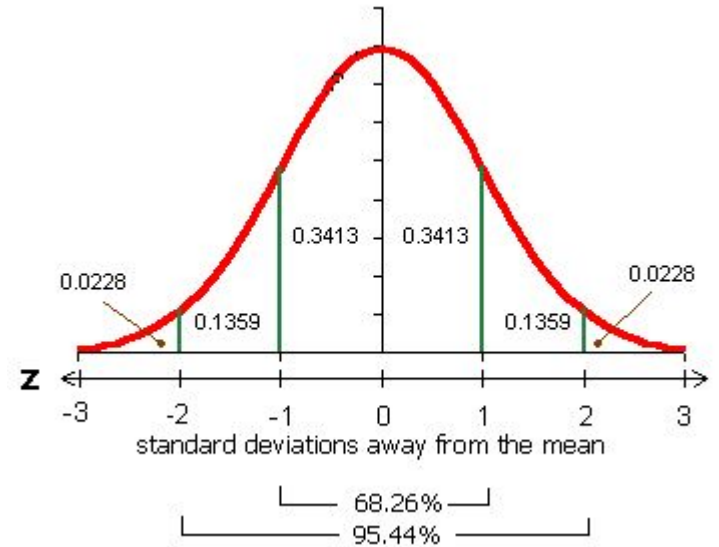
Ejemplo

Imaginemos una clase. Estas son las notas de un examen.

6	6	8	9	9
9	6	8	10	8
5	4	9	8	9
3	6	6	7	6
10	3	4	5	4

Normalizemos

-0.34	-0.34	0.60	1.07	1.07
1.07	-0.34	0.60	1.54	0.60
-0.81	-1.28	1.07	0.60	1.07
-1.75	-0.34	-0.34	0.13	-0.34
1.54	-1.75	-1.28	-0.81	-1.28



$$z = \frac{x - \mu}{\sigma}$$

Moran Local LISA

Si bien Moran Global nace para analizar autocorrelación espacial, su variante local puede utilizarse para detectar cluster espaciales. Estos se denominan LISA o Local Indicators of Spatial Association.

Calculado para celda del medio (42) con
filas estandarizadas y pesos torre (rook)

45 44 44

43 42 39

38 32 34

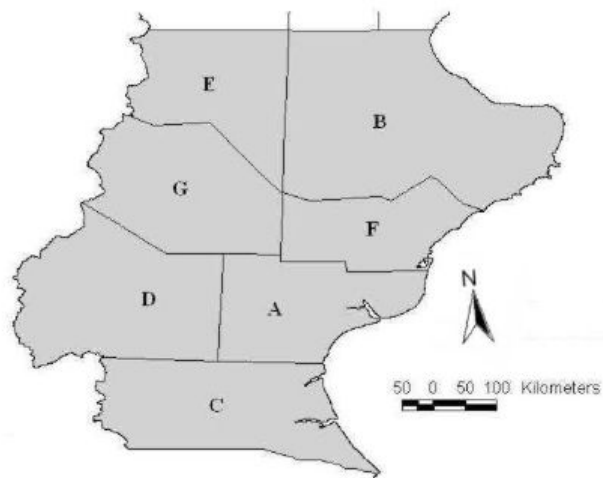
$$I_i = \left[\frac{z_i}{s^2} \right] \sum_{j=1}^n w_{ij} z_j, j \neq i$$

y_i	z_i	w_{ij}	$w_{ij}z_j$
45	4.889	0.000	0.000
43	2.889	0.250	0.722
38	-2.111	0.000	0.000
44	3.889	0.250	0.972
42	1.889	0.000	0.000
32	-8.111	0.250	-2.028
44	3.889	0.000	0.000
39	-1.111	0.250	-0.278
34	-6.111	0.000	0.000
Sum		1.000	-0.611

Otra posible formulacion

$$I_i = \frac{(P_{xi} - \bar{P}_x) \sum_j c_{ij} (P_{xj} - \bar{P}_x)}{\sum_i \frac{(P_{xi} - \bar{P}_x)^2}{N}}$$

Porcentaje de viviendas con baño exclusivo por departamento. Pcia. De Santa Cruz



<i>Departamento</i>	<i>% Baño Exclusivo</i>
A) Corpen Aike	98,92
B) Deseado	95,36
C) Guer Aike	96,35
D) Lago Argentino	97,26
E) Lago Buenos Aires	98,21
F) Magallanes	98,33
G) Rio Chico	96,15
<i>Promedio</i>	<i>97,23</i>
<i>Varianza</i>	<i>10,46</i>

Fuente: Celemin
(2000)

1) establecer la matriz de pesos a partir de la cantidad de vecinos que tiene cada departamento (contigüidad tipo Rook)

Matriz de pesos

	A	B	C	D	E	F	G	Total
A	0	0	1	1	0	1	1	4
B	0	0	0	0	1	1	0	2
C	1	0	0	1	0	0	0	2
D	1	0	1	0	0	0	1	3
E	0	1	0	0	0	0	1	2
F	1	1	0	0	0	0	1	3
G	1	0	0	1	1	1	0	4
Suma								20

Matriz de pesos estandarizada

	A	B	C	D	E	F	G	Total
A	0	0	0,25	0,25	0	0,25	0,25	1
B	0	0	0	0	0,5	0,5	0	1
C	0,5	0	0	0,5	0	0	0	1
D	0,33	0	0,33	0	0	0	0,33	1
E	0	0,5	0	0	0	0	0,5	1
F	0,33	0,33	0	0	0	0	0,33	1
G	0,25	0	0	0,25	0,25	0,25	0	1

Departamento	% Baño Exclusivo (x_i)	Promedio (\bar{x})	$x_i - \bar{x}$ (z_i)	Varianza (m_2)	z_i / m_2
A) Corpen Aike	98,92	97,23	1,69	10,46	0,1616
B) Deseado	95,36	97,23	-1,87	10,46	-0,1788
C) Guer Aike	96,35	97,23	-0,88	10,46	-0,0841
D) Lago Argentino	97,26	97,23	0,03	10,46	0,0029
E) Lago Buenos Aires	98,21	97,23	0,98	10,46	0,0937
F) Magallanes	98,33	97,23	1,10	10,46	0,1051
G) Río Chico	96,15	97,23	-1,08	10,46	-0,1032

$$I_A = (98,92 - 97,23) / 10,46 \times [(0,25 \times -0,88) + (0,25 \times 0,03) + (0,25 \times 1,10) + (0,25 \times -1,08)] \\ 0,1616 \times [-0,22 + 0,0075 + 0,275 + (-0,27)] = -0,0335$$

$$I_B = (95,36 - 97,23) / 10,46 \times [(0,5 \times 0,98) + (0,5 \times 1,10)] \\ -0,1788 \times [0,49 + 0,55] = -0,1859$$

$$I_C = (96,35 - 97,23) / 10,46 \times [(0,5 \times 1,69) + (0,5 \times 0,03)] \\ -0,0841 \times [0,845 + 0,015] = -0,0723$$

$$I_D = (97,26 - 97,23) / 10,46 \times [(0,33 \times 1,69) + (0,33 \times -0,88) + (0,33 \times -1,08)] \\ 0,0029 \times [0,5577 + (-0,2904) + (-0,3564)] = -0,0003$$

$$I_E = (98,21 - 97,23) / 10,46 \times [(0,5 \times -1,87) + (0,5 \times -1,08)] \\ 0,0937 \times [-0,935 + (-0,54)] = -0,1382$$

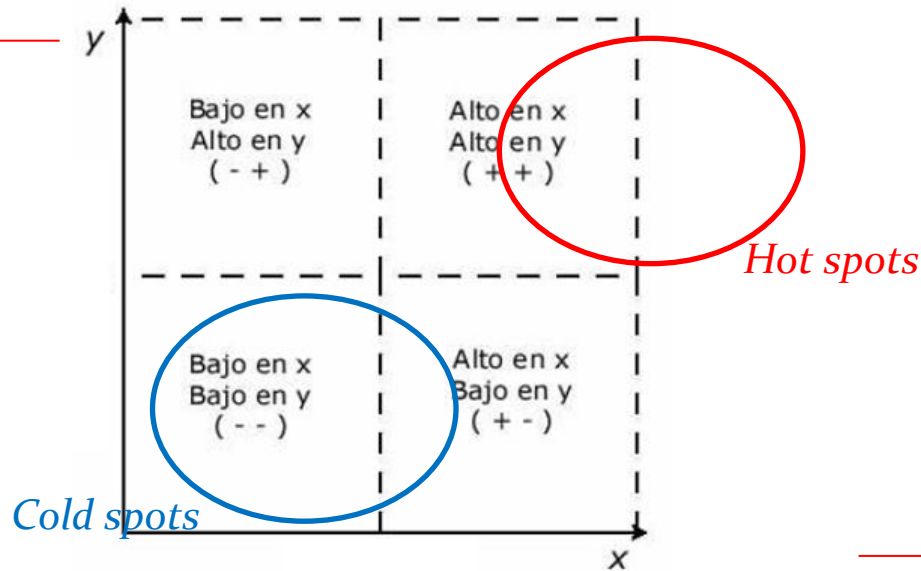
$$I_F = (98,33 - 97,23) / 10,46 \times [(0,33 \times 1,69) + (0,33 \times -1,87) + (0,33 \times -1,08)] \\ 0,1051 \times [0,5577 + (-0,6171) + (-0,3564)] = -0,0437$$

$$I_G = (96,15 - 97,23) / 10,46 \times [(0,25 \times 1,69) + (0,25 \times 0,03) + (0,25 \times 0,98) + (0,25 \times 1,10)] \\ -0,1032 \times [0,4225 + 0,0075 + 0,245 + 0,275] = -0,0980$$

La suma de los valores locales ($I_A + I_B + I_C + I_D + I_E + I_F + I_G$) conforman el valor global de I de Moran = 0,5720

El *Moran Scatterplot* es una herramienta de análisis que permite en un gráfico de dispersión el comportamiento de cada unidad espacial

Valores estandarizados
del promedio de los
valores en unidades
vecinas



Valores estandarizados
de una variable para
cada unidad espacial

Fuente: Buzai (2005)

Getis Ord – Gi*

Aplicaciones de Mapas de Puntos Cálidos-Fríos

✓ **Epidemiology** [e.g: Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181), 990-993.]

✓ **Criminology** [e.g: Chakravorty, S. (1995). Identifying crime clusters: The spatial principles. *Middle States Geographer*, 28, 53-58.]

✓ **Ecology** [e.g: Mahboubi, P., Parkes, M., Stephen, C., & Chan, H. M. (2015). Using expert informed GIS to locate important marine social-ecological hotspots. *Journal of environmental management*, 160, 342-352.]

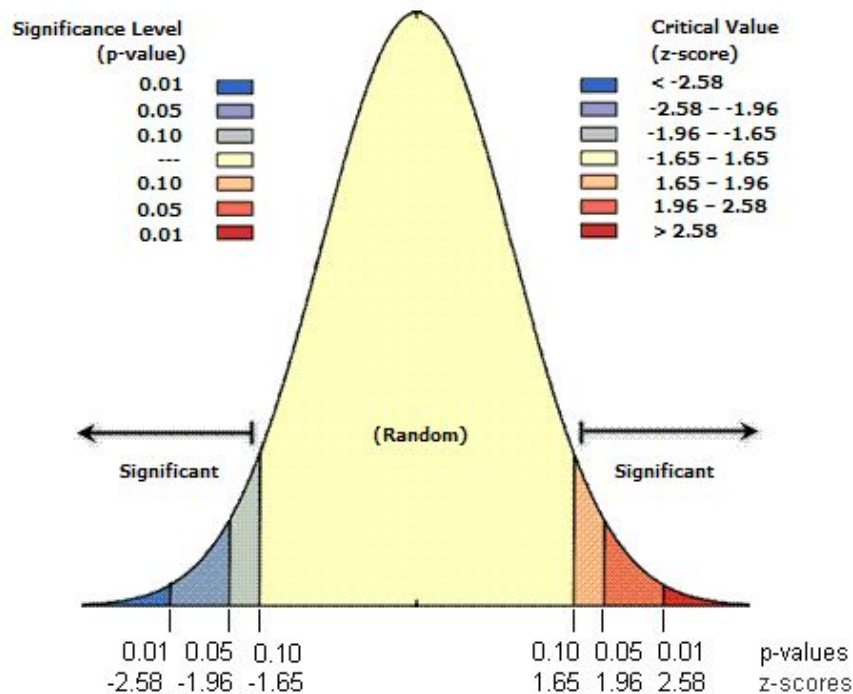
✓ **Transportation Management** [e.g: Cheng, W., & Washington, S. P. (2005). Experimental evaluation of hotspot identification methods. *Accident Analysis and Prevention*, 37(5), 870-881.]

✓ **User Generated Content Analysis** [e.g: Brovelli, M. A., Oxoli, D., & Zurbarán, M. A. (2016). Sensing Slow Mobility and Interesting Locations for Lombardy Region (Italy): a Case Study Using Pointwise Geolocated Open Data. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 603-607.]

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

σ = Standard Deviation



Test de Hipotesis

	Naturaleza de la hipótesis nula	
Decisión sobre H_0	H_0 cierta	H_0 Falsa
Aceptar H_0	Correcto	Error de tipo II Probabilidad β
Rechazar H_0 Acepto H_1	Error de tipo I Probabilidad α	Correcto

Getis Ord Gi*

Las puntuaciones z y los valores P resultantes indican dónde se agrupan espacialmente las entidades con valores altos o bajos. Esta técnica funciona mediante la búsqueda de cada entidad dentro del contexto de entidades vecinas, definidas por la matriz w . Una entidad con un valor alto es interesante, pero es posible que no sea un punto caliente estadísticamente significativo. Para ser un punto caliente estadísticamente significativo, una entidad debe tener un valor alto y también estar rodeada por otras entidades con valores altos.

Formula para el cálculo

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j} \right)^2}{n-1}}} \quad (1)$$

where x_j is the attribute value for feature j , $w_{i,j}$ is the spatial weight between feature i and j , n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (3)$$

The G_i^* statistic is a z -score so no further calculations are required.

Getis Ord G_i^*

La suma local para una entidad y sus vecinos se compara proporcionalmente con la suma de todas las entidades; cuando la suma local es muy diferente de la esperada, y esa diferencia es demasiado grande como para ser el resultado de una opción aleatoria, se obtiene como consecuencia una puntuación z estadísticamente significativa.

HotSpot Analysis Plugin

https://github.com/danioxoli/HotSpotAnalysis_Plugin

Installation - Windows

1) Install dependencies:

Open `osGeo4W Shell` installed with QGIS3 as `Administrator` and type:

```
$ py3_env.bat
$ python3 -m pip install --upgrade pip
$ python3 -m pip install pysal==1.14.3 --user
```

This will install in your QGIS3 Python 3 environment all the required dependencies. Be sure that the installed PySAL version is $\leq 1.14.3$. Run `python -m pip show pysal` for checking it. If the installed version is higher, type `python -m pip install -I pysal==1.14.3`

2) Open QGIS:

Go to `Plugins -> Manage and Install plugins -> Settings -> Show also experimental plugins`

In `All plugins` tab, look for `Hotspot Analysis` and tick the checkbox.

A new icon for Hotspot Analysis will appear on the QGIS main panel and in the Vector Menu.

Python



Python es un lenguaje de programación interpretado cuya filosofía hace hincapié en una sintaxis que favorezca un código legible. Fue desarrollado por Guido van Rossum

Se trata de un lenguaje de programación multiparadigma, ya que soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional.

Actualmente tiene dos versiones 2.7 y 3

Su licencia es libre y compatible con GPL (GNU Public License)

El Zen de Python

- Lindo es mejor que feo.
- **Explícito es mejor que implícito.**
- Simple es mejor que complejo.
- Complejo es mejor que complicado.
- Plano es mejor que anidado.
- Espaciado es mejor que denso.
- La legibilidad es importante.
- Los casos especiales no son lo suficientemente especiales como para romper las reglas.
- **Sin embargo la practicidad le gana a la pureza.**

El Zen de Python (por Tim Peters)

- **Los errores nunca deberían pasar silenciosamente.**
- A menos que se silencien explícitamente.
- **Frente a la ambigüedad, evitá la tentación de adivinar.**
- **Debería haber una, y solamente una, manera obvia de hacerlo.**
- A pesar de que no sea obvio a menos que seas Holandés (como GvR)
- **Ahora es mejor que nunca.**
- **A pesar de que nunca es muchas veces mejor que *justo ahora*.**
- Si la implementación es difícil de explicar, es una mala idea.
- **Si la implementación es fácil de explicar, quizás sea una buena idea.**
- Los espacios de nombres son una gran idea, ¡tengamos más de esos!

Librerías de Python

Las librerías de Python pueden considerarse como plug-ins, funcionalidades adicionales desarrolladas por la comunidad que le suman herramientas al programa original.

Por eso la importancia de usar software libre, puede canalizarse la energía de una comunidad dispersa a nivel mundial con enorme potencial

- `Import this`
- `Import numpy`

Numpy



NumPy es una extensión de Python, que le agrega mayor soporte para vectores y matrices, constituyendo una biblioteca de funciones matemáticas de alto nivel para operar con esos vectores o matrices. El ancestro de NumPy, Numeric, fue creado originalmente por Jim Hugunin con algunas contribuciones de otros desarrolladores.

En 2005, Travis Oliphant creó NumPy incorporando características de Numarray en NumPy con algunas modificaciones. NumPy es software **LIBRE**

Scipy

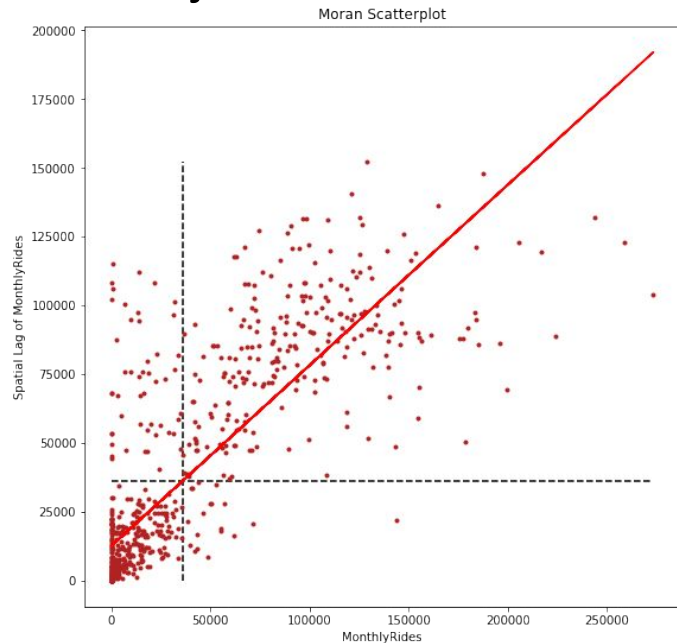


SciPy es una biblioteca **LIBRE** de herramientas y algoritmos matemáticos para Python que nació a partir de la colección original de Travis Oliphant que consistía de módulos de extensión para Python, lanzada en 1999.

SciPy contiene módulos para optimización, álgebra lineal, integración, interpolación, funciones especiales, FFT, procesamiento de señales y de imagen, resolución de ODEs y otras tareas para la ciencia e ingeniería.

Pysal

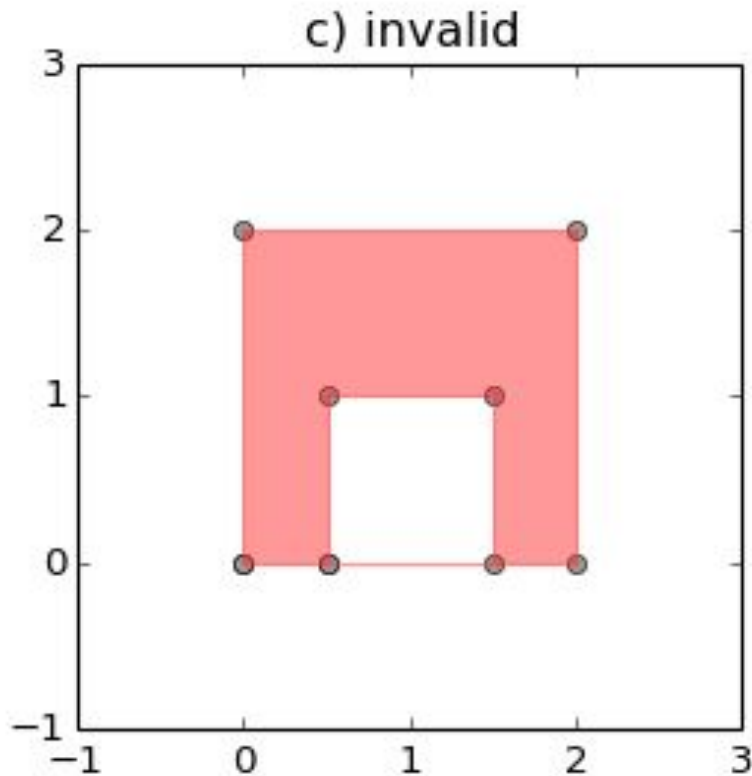
PySAL es una biblioteca de código abierto de funciones de análisis espacial escrita en Python destinada a apoyar el desarrollo de aplicaciones de alto nivel. PySAL es de código abierto bajo la licencia **LIBRE**.



Shapely

Es una librería que permite la creación, edición y manipulación de objetos espaciales.

```
ext = [(0, 0), (0, 2), (2, 2), (2, 0), (0, 0)]  
int = [(0.5, 0), (1.5, 0), (1.5, 1), (0.5, 1), (0.5, 0)]  
polygon = Polygon(ext, [int])
```



Geopandas

El objetivo de GeoPandas es facilitar el trabajo con datos geoespaciales en Python. Combina las capacidades de los pandas y las formas, proporcionando operaciones geoespaciales en pandas y una interfaz de alto nivel para múltiples geometrías a formas. GeoPandas le permite realizar fácilmente operaciones en Python que de otra manera requerirían una base de datos espacial como PostGIS.

```
# Plot by GDP per capita  
In [5]: world = world[(world.pop_est>0) & (world.name!="Antarctica")]  
  
In [6]: world['gdp_per_cap'] = world.gdp_md_est / world.pop_est  
  
In [7]: world.plot(column='gdp_per_cap');
```

