



Análise Exploratória de Dados Espaciais com Python

Gesiel Rios Lopes¹, Karina Jorge Pelarigo¹,
Alexandre C. B. Delbem¹, Joélcio Braga de Sousa¹

¹Laboratório de Sistemas Embarcados e Evolutivos (LSEE)

Instituto de Ciências Matemática e de Computação (ICMC)

Universidade de São Paulo (USP)

gesielrios@usp.br, acbd@icmc.usp.br

Agenda

- Introdução à análise exploratória de dados espaciais
 - Por que usar python para análise exploratória de dados?
- Dados espaciais
- Matrizes de ponderação espacial
- Peso de contiguidade
- Mapas coropléticos
 - Classificação de dados
 - Comparando esquemas de classificação
- Autocorrelação espacial global
 - Estatística I de Moran
 - Outros índices globais
- Autocorrelação espacial local
 - I de Moran local
- Estatística local de Getis e Ord

About me!

Gesiel Rios Lopes



ESALQ



UNIFOR
ENSINANDO E APRENDENDO





1.

Introdução à análise exploratória de dados espaciais



Introdução

- ◎ Análise Exploratória de Dados Espaciais - AEDE
 - *Exploratory Spatial Data Analysis ESDA*
 - “Coleção de técnicas para descrever e visualizar distribuições espaciais, identificar situações atípicas, descobrir padrões de associação espacial, clusters e sugerir regimes espaciais ou formas de heterogeneidade espacial” (Anselin).

Introdução

◎ Análise Exploratória de Dados Espaciais - AEDE

1. Visualização de padrão de áreas

- ◎ técnicas convencionais de visualização cartográfica, estatísticas não-espaciais.

2. Indicadores Globais de Autocorrelação

- ◎ explorar a dependência espacial, mostrando como os valores estão correlacionados no espaço;
- ◎ o conceito utilizado é o de autocorrelação espacial;
- ◎ exemplos de indicadores globais: Moran (I), Geary (C), Getis e Ord (G).

Introdução



◎ Análise Exploratória de Dados Espaciais - AEDE

3. Indicadores Locais de Associação Espacial (LISA)

- ◎ Possibilita a identificação de:
 - clusters: objetos com valores de atributos semelhantes;
 - outliers: objetos anômalos;
 - a presença de mais de um regime espacial.
- ◎ exemplos de indicadores locais: Moran (I_i), Getis e Ord (G_i e G_i^*).



1.2.

Por que usar python?

O que é Python ?

- Uma linguagem de programação de alto nível (*Very High Level Language*).
- Multiplataforma
 - Unix, Windows, Symbian, Solaris, etc...
- Multiparadigma
 - Orientada a objetos, funcional, orientação a aspectos, programação procedural.
- A tipagem dos dados e objetos é dinâmica e forte.
- Interpretada ou Compilada.
- Open Source (Free Software - GLP).



Por que Python ?

- Fácil aprendizado
- Sintaxe limpa e de fácil leitura
- Forte suporte da comunidade
- Forte documentação
- Biblioteca padrão enorme
- Divertida
- Mais com menos [códigos]
- Liberdade



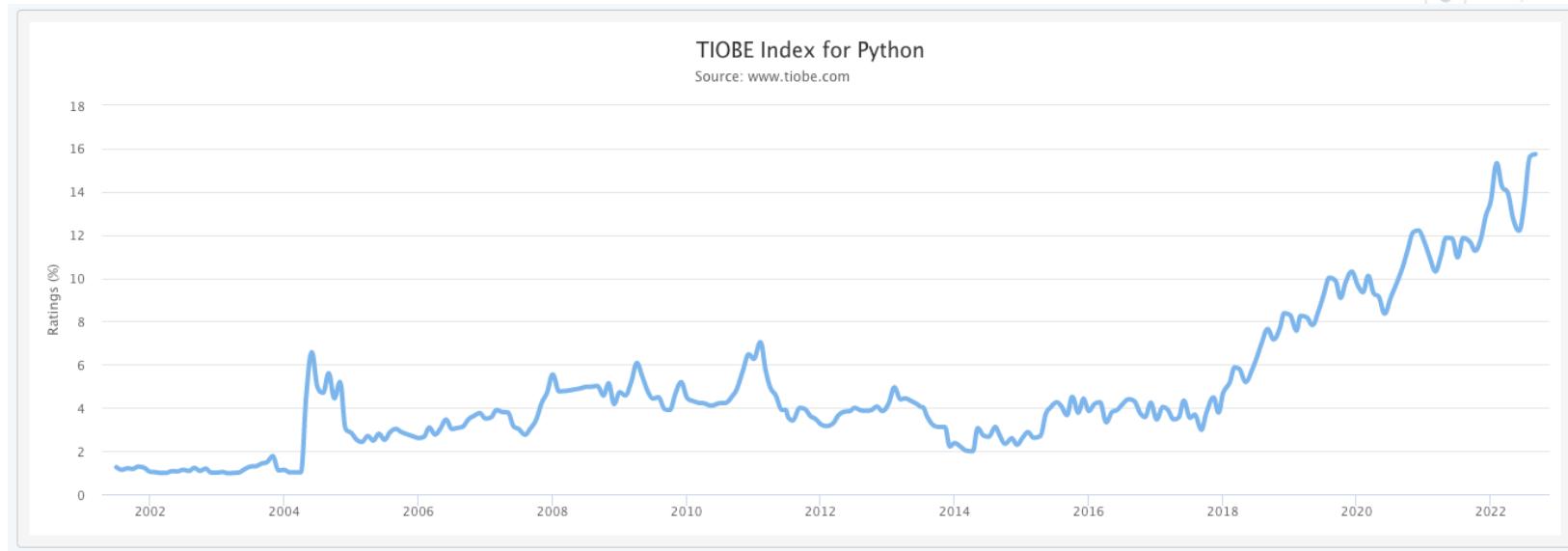
Por que Python ?

Sep 2022	Sep 2021	Change	Programming Language	Ratings	Change
1	2	▲	 Python	15.74%	+4.07%
2	1	▼	 C	13.96%	+2.13%
3	3		 Java	11.72%	+0.60%
4	4		 C++	9.76%	+2.63%
5	5		 C#	4.88%	-0.89%
6	6		 Visual Basic	4.39%	-0.22%
7	7		 JavaScript	2.82%	+0.27%
8	8		 Assembly language	2.49%	+0.07%
9	10	▲	 SQL	2.01%	+0.21%
10	9	▼	 PHP	1.68%	-0.17%

Fonte: <https://www.tiobe.com/tiobe-index/>, acesso em 28/09/2022

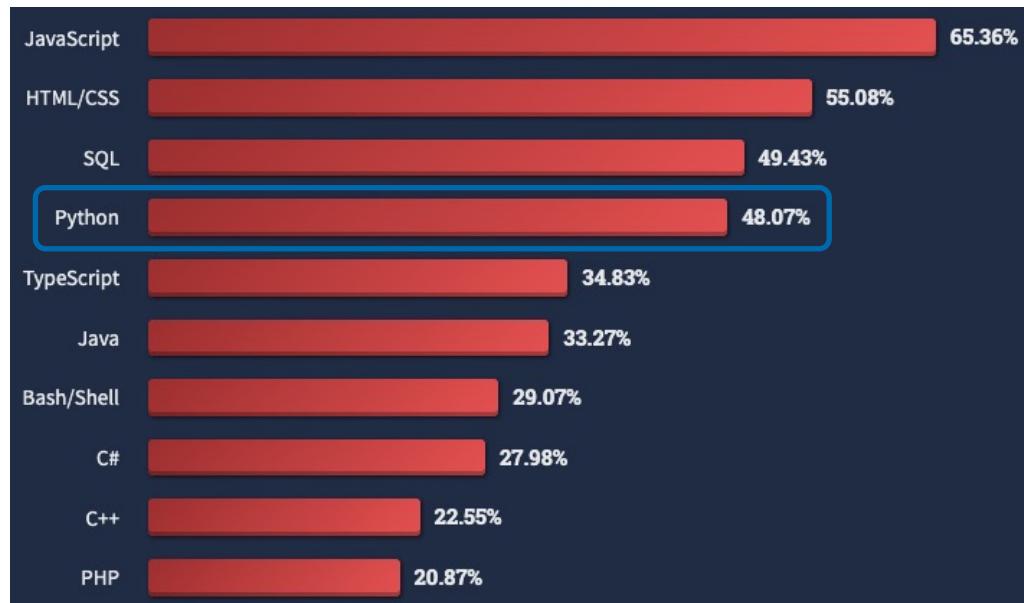


Por que Python ?



Fonte: <https://www.tiobe.com/tiobe-index/python/>, acesso em 28/09/2022.

Por que Python ?



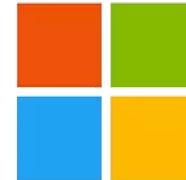
Fonte: <https://survey.stackoverflow.co/2022/#technology-most-popular-technologies>, acesso em 28/09/2022.



Quem usa Python ?



Google



Microsoft



The
WALT DISNEY
Company



ubuntu



Red Hat



Massachusetts
Institute of
Technology

Por onde começar ?

- ◎ Tutoriais e cursos:

- <https://docs.python.org/3/tutorial/index.html>
- <https://www.pycursos.com/python-para-zumbis/>
- <https://www.datascienceacademy.com.br/course?courseid=python-fundamentos>
- <https://www.udemy.com/course/python-3-completo/>
- <https://www.alura.com.br/cursos-online-programacao/python>



Por onde começar ?

- ◎ No desenvolvimento de aplicações científicas é recomendável utilizar a distribuição Anaconda¹, uma vez que ela é de código aberto e é a maneira mais fácil de executar aplicações científicas desenvolvidas em Python no Linux, Windows e Mac OS X.



¹Disponível em: <https://www.anaconda.com/distribution/>, acesso em 16/07/2021.

E agora ?

- ◎ Para a conclusão do minicurso iremos utilizar o jupyter notebook;
- ◎ Permite a execução de códigos em células, permitindo um bom uso em tarefas de análise de dados.



Ferramentas



Shapely Fiona PyProj



matplotlib seaborn

pandas GeoPandas



2. **Dados espaciais**

Dados espaciais

- ◎ Dados espaciais, dados geográficos ou dados geoespaciais têm duas características muito úteis.
 - Primeiro, os dados geográficos são onipresentes.
 - Tudo tem uma localização no espaço-tempo e essa localização pode ser usada diretamente para fazer melhores previsões ou inferências.

Dados espaciais

- ◎ Dados espaciais podem ser definidos como dados ou informações associados às coordenadas geográficas que representam a superfície terrestre.
- ◎ Os tipos mais comuns de dados espaciais são:
 - dados *raster* ou matriciais; e
 - dados vetoriais.

Dados espaciais

◎ dados *raster* ou matriciais

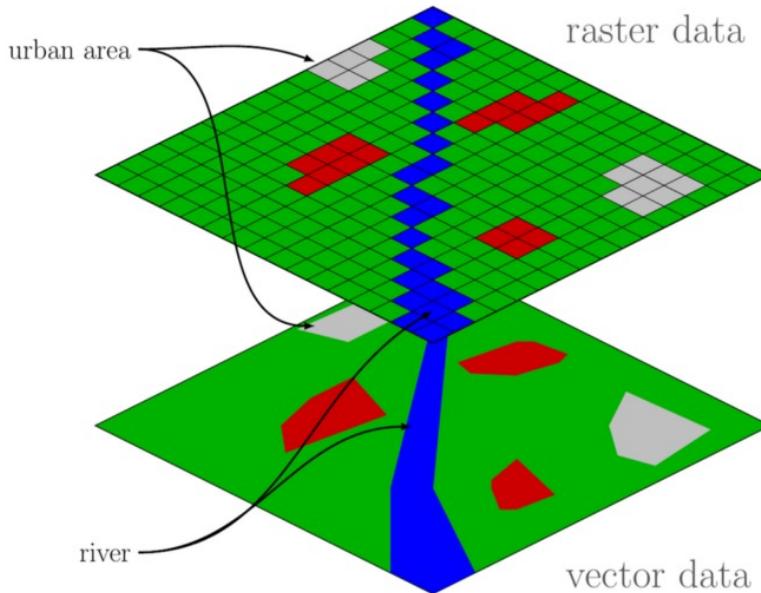
- Os dados em formato *raster* estão em uma tabela (linhas e colunas) onde cada célula (também chamada de *pixels*) contém informações sobre a área que representa.
- Na maioria das vezes, as células são quadradas e regularmente espaçadas.
- As informações em cada célula podem ser uma cor, uma altitude, uma temperatura ou qualquer outro indicador que o autor do arquivo queira transmitir.
dados vetoriais.

Dados espaciais

◎ dados vetoriais

- representam entidades geográficas a partir de três elementos básicos: pontos, linhas e polígonos (áreas).
- Por sua vez, cada entidade/objeto geográfico pode armazenar elementos adicionais.
- Por exemplo, se cada entidade for um país, o nome do país, o tamanho da população, o idioma oficial etc. poderão ser armazenados.

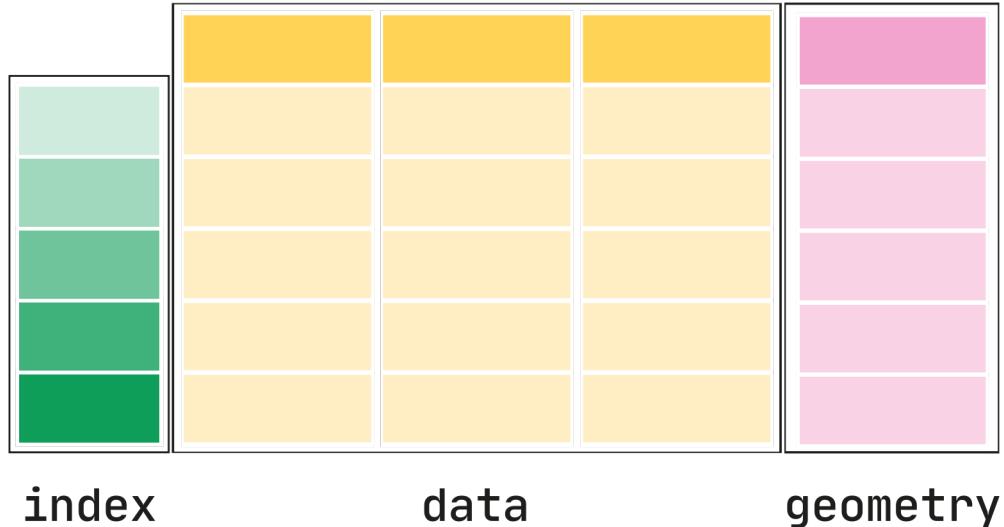
Dados espaciais



Fonte: <https://blog.rmotr.com/spatial-data-with-python-lets-begin-e29b5c41ead3>, acesso em 28/09/2022

Dados espaciais

◎ Tabelas geográficas





3. **Matrizes de Ponderação Espacial**

Matrizes de Ponderação Espacial

- ◎ Tem dimensão $n \times n$, em que n é o número de áreas (objetos).
- ◎ Os elementos w_{ij} representam uma medida de proximidade entre objetos O_i e O_j .

Matrizes de Ponderação Espacial

◎ Pesos de contiguidade

$$w_{ij} = \begin{cases} 1 & \text{se } i \text{ e } j \text{ são contíguos} \\ 0 & \text{se } i \text{ e } j \text{ não são contíguos} \end{cases}$$

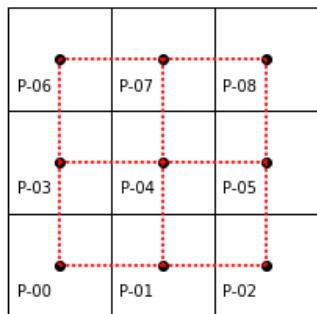


	A	B	C	D	E
A	0	1	0	1	0
B	1	0	1	1	1
C	0	1	0	0	1
D	1	1	0	0	1
E	0	1	1	1	0

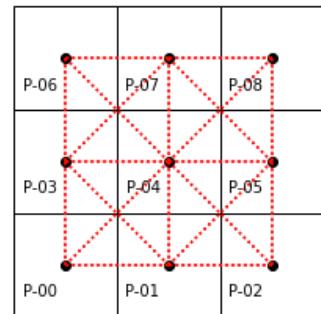
Matrizes de Ponderação Espacial

◎ Pesos de contiguidade

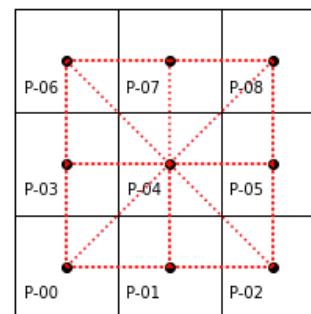
ROOK



QUEEN



KNN - $k = 3$



4.

Mapas coropléticos

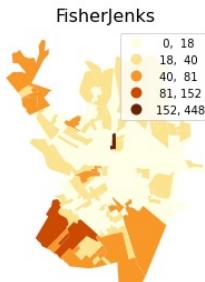
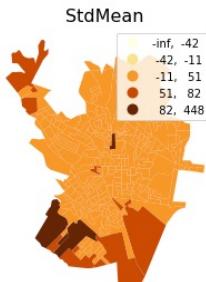
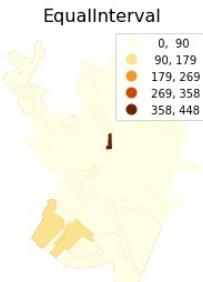
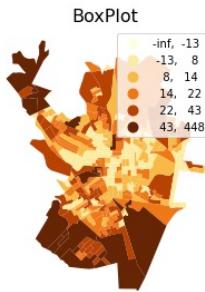
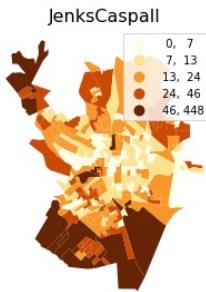
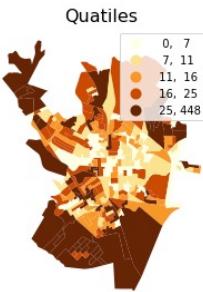
Mapas coropléticos

- ◎ Mapas coropléticos são mapas geográficos que exibem informações estatísticas codificadas em uma paleta de cores.
- ◎ Forma usual de apresentação de dados agregados por áreas.
- ◎ A eficácia de um mapa coroplético depende na maioria do propósito do mapa.

Qual mensagem você deseja comunicar moldará quais opções são preferíveis em relação a outras.

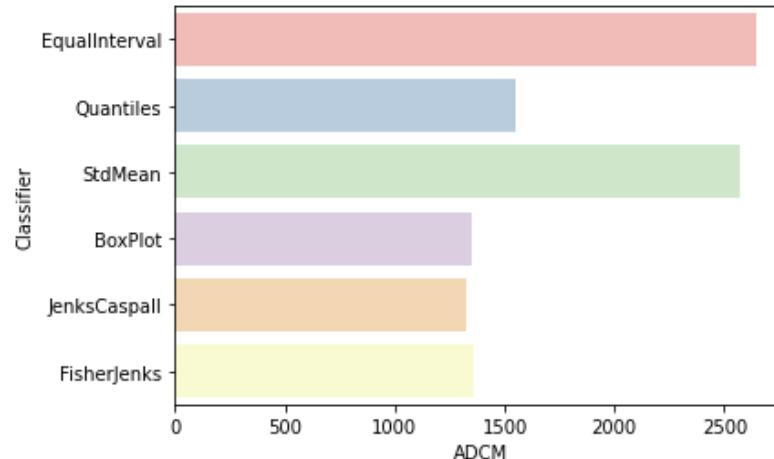
Mapas coropléticos

◎ Visualização de padrões de área



Mapas coropléticos

- Comparando esquemas de classificação
 - desvio absoluto em torno das medianas de classe (do inglês *absolute deviation around class medians*





4. Autocorrelação Espacial Global

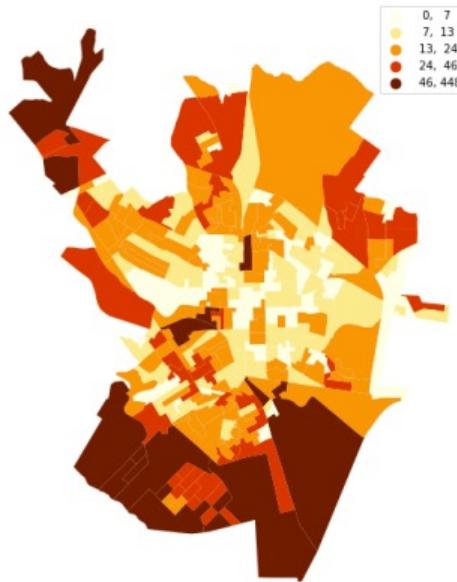
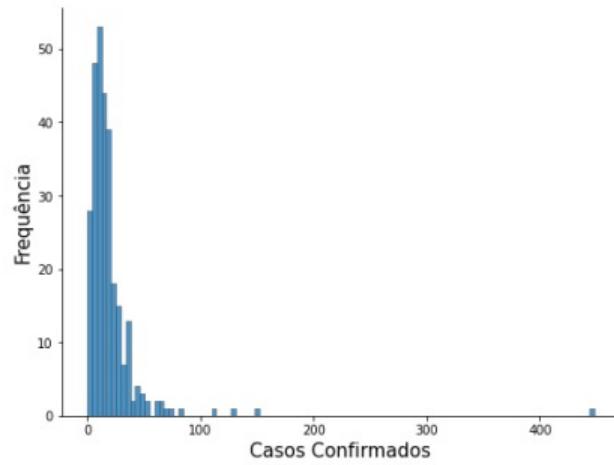
Autocorrelação Espacial Global

- ◎ A noção de autocorrelação espacial refere-se à existência de uma “relação funcional entre o que acontece em um ponto do espaço e o que acontece em outro”.
 - tem a ver com o grau onde a similaridade de valores entre observações em um conjunto de dados está relacionada à similaridade nas localizações de tais observações.

Autocorrelação Espacial Global

- ◎ Ao analisar um estamos buscando padrões do ponto de vista espacial;
 - um dos aspectos necessários é verificar se o evento em estudo e os fatores relacionados a ele possuem distribuição espacialmente condicionada (dependência espacial), ou seja, se os valores do atributo em estudo em uma determinada região depende ou não dos valores deste atributo nas regiões vizinhas.

Autocorrelação Espacial Global



Autocorrelação Espacial Global

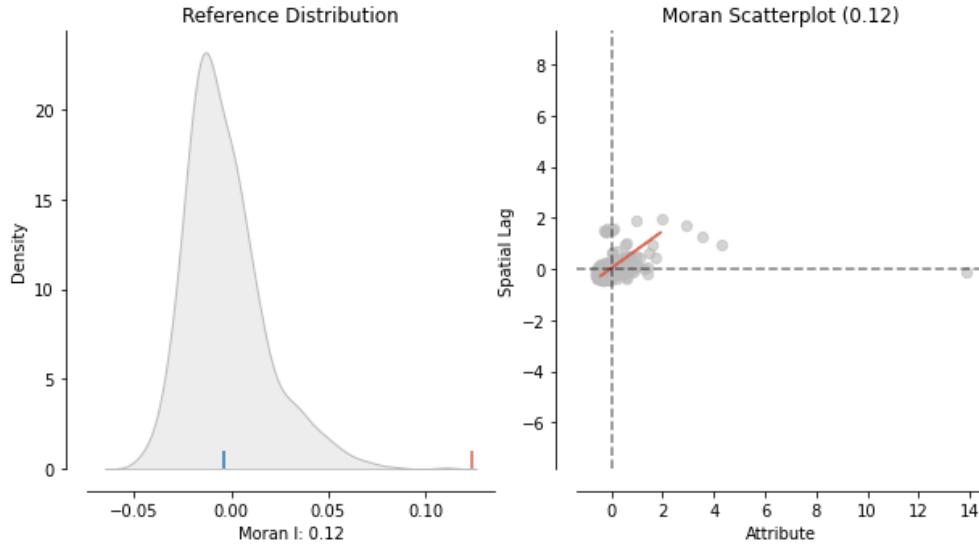
◎ Índice / de Moran

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}$$

em que n é o número de observações, z_i é o valor padronizado da variável de interesse na região i , e w_{ij} é a célula correspondente à i -ésima linha da j -ésima coluna da matriz de pesos espaciais.

Autocorrelação Espacial Global

Índice / de Moran



Autocorrelação Espacial Global

- ◎ Indicadores globais como o de Moran fornecem um valor único como medida da associação espacial para todo o conjunto de dados, o que é útil na caracterização da região como um todo.

Autocorrelação Espacial Global

- ◎ Outros índices globais

- Índice C de Geary

$$C = \frac{(n-1)}{2\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_i - y_j)^2}{\sum_i (y_i - \bar{y})^2}$$

- Índice G de Getis e Ord

$$G(d) = \frac{\sum_i \sum_j w_{ij}(d) y_i y_j}{\sum_i \sum_j y_i y_j}$$



4. **Autocorrelação Espacial Local**

Autocorrelação Espacial Local

- ◎ Indicadores globais de Moran e Geary fornecem um valor único como medida da associação espacial.
 - Muitas vezes é necessário examinar padrões numa escala local.
 - Neste caso, é preciso utilizar indicadores locais de associação espacial que possam ser associados a diferentes localizações de uma variável distribuída espacialmente.
 - A utilização dos indicadores locais em conjunto com os indicadores globais refinam nosso conhecimento sobre o processos que dão origem a dependência espacial.

Autocorrelação Espacial Local

- ◎ Os indicadores locais de associação espacial produzem um valor específico para cada objeto (área).
- ◎ Isto acarreta a identificação de:
 - clusters: objetos com valores de atributos semelhantes;
 - outliers: objetos anômalos;
 - a presença de mais de um regime espacial.

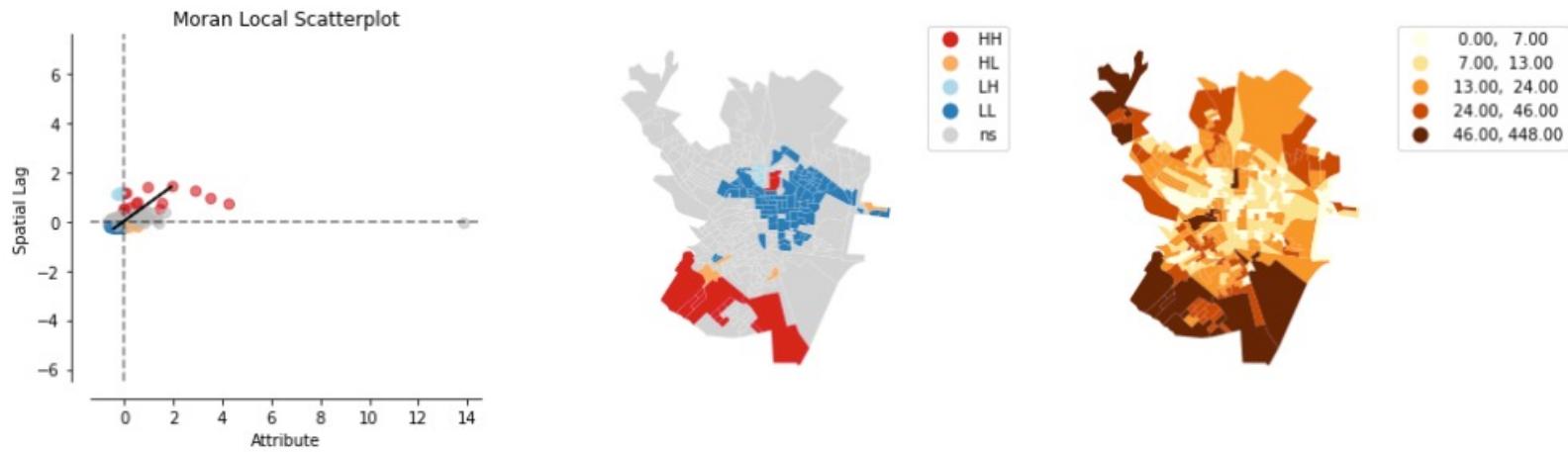
Autocorrelação Espacial Local

◎ I de Moran local

- Indicadores Locais de Associação Espacial (do inglês, *Local Indicators of Spatial Association* - LISA)
- visa identificar casos em que o valor de uma observação e a média de seus arredores são:
 - semelhantes (alto-alto, do inglês high-high - HH ou baixo-baixo, do inglês low-low - LL no gráfico de dispersão de Moran)
 - diferentes (alto-baixo, do inglês high-low - HL ou baixo-alto, do inglês low-high - LH)

Autocorrelação Espacial Local

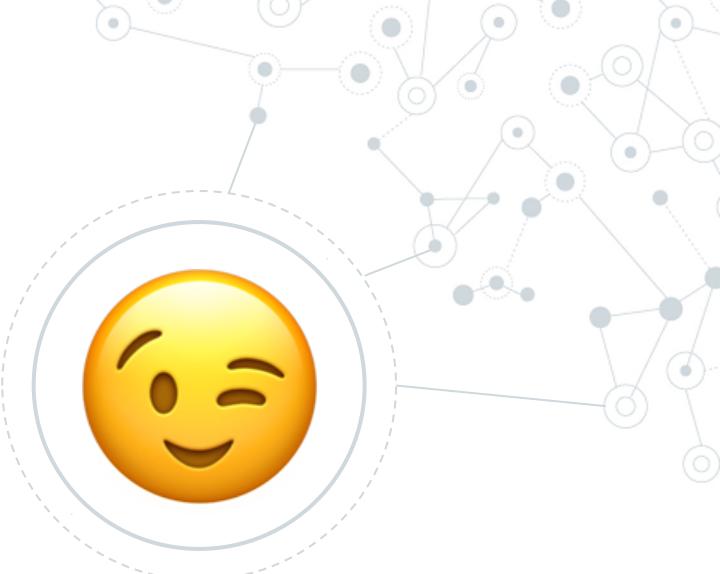
◎ / de Moran local





Obrigado!

Perguntas ?



github.com/gesielrios



gesielrios@usp.br



Agradecimentos



CAPES

Conselho Nacional de Desenvolvimento
Científico e Tecnológico

Fundação de Amparo à Pesquisa e ao Desenvolvimento
Científico e Tecnológico do Maranhão



Center for
Artificial
Intelligence

LSEE



Laboratório de Sistemas Embarcados e Evolutivos



ERCERMAPI
Escola Regional de Computação
do Ceará, Maranhão e Piauí



PARTNER
Trusted Advisors



That's all Folks!