

Econometria Espacial
Programa de Pós-graduação em Economia Aplicada
Universidade Federal de Juiz de Fora

Professor: Eduardo Almeida

Monitor: Tauã Magalhães VitalVital

Análise Exploratória de Dados Espaciais
Introdução ao GEODA

Geoda é um software livre desenvolvido por Luc Anselin e seu time em 2003. Desde então o software vem sendo utilizado para Análise Exploratória de Dados Espaciais (AEDE) e análises mais simples de dados espaciais. De acordo com seus fundadores, o software foi desenvolvido com o intuito de facilitar a exploração e análise de dados espaciais. O software não exige GIS (*Geographic Information System*), mas é compatível com essa estrutura de dados.

O download da versão mais recente do software pode ser feito no link <http://geodacenter.github.io/download.html>. Atente-se para a escolha do sistema operacional.

GeoDa Release

Help us keep GeoDa free by contributing [here](#). Thank you for supporting free and open-source spatial software!

GeoDa is continuously updated. The most current version is GeoDa 1.16 with [new features](#). We found it to be stable but if you encounter a bug, [please let us know](#). You can also report bugs and issues on [our Github site](#). The source code for GeoDa is available [here](#) and is released under the [GPL v3.0 license](#).



[Download For Windows](#)

[Download For Mac](#)

[Download For Linux](#)

[Nightly build](#)

Figura 1: Download GEODA software

Na versão mais recente, tem-se a opção de utilizar diversos formatos de dados (shp, csv, SQL, entre outros).

1. Carregando dados

Iremos iniciar o uso do GEODA carregando dados areais sobre AirBnB em Chicago. Os dados bem como a descrição detalhada dos mesmos estão disponíveis em <https://geodacenter.github.io/data-and-lab/airbnb/>.

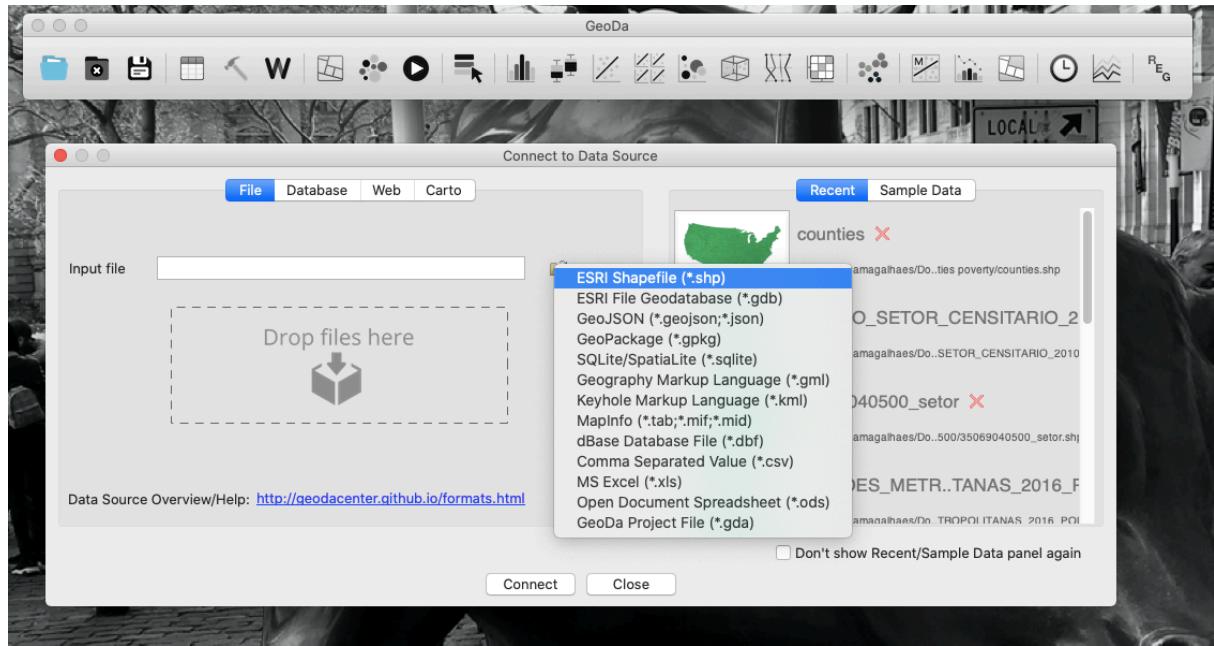


Figura 2: carregando dados.

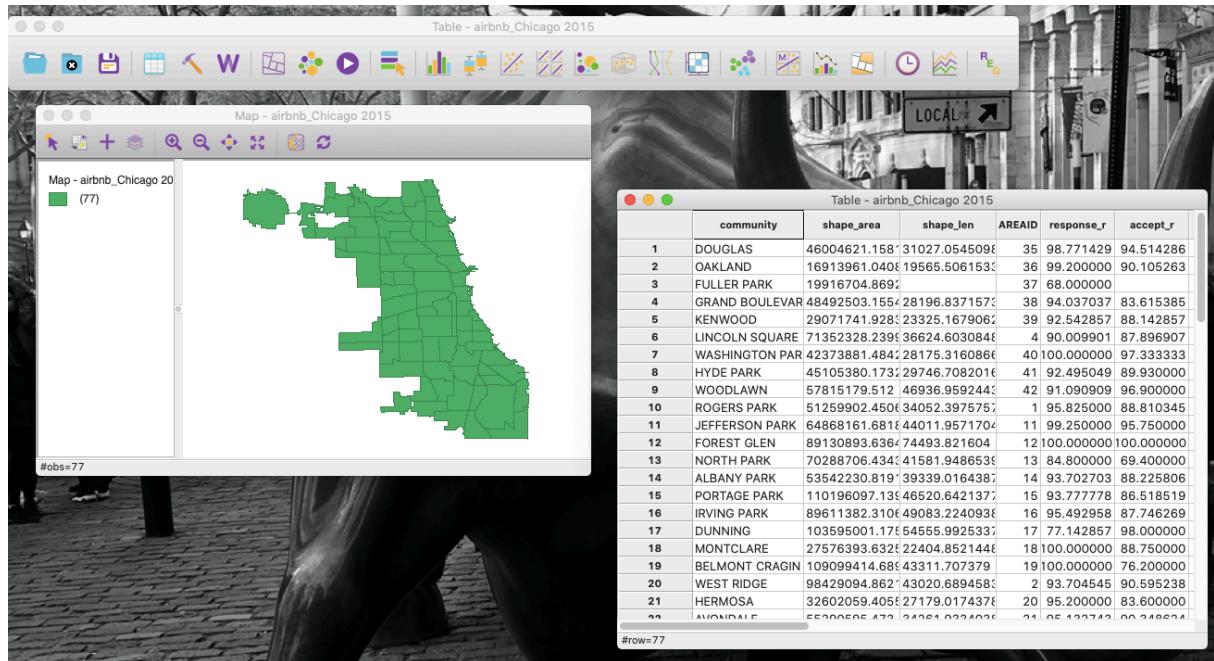


Figura 3: Dados AirBnB

2. Merge de dados externos

A barra de opções acima fornece a opção de unir um conjunto de dados externos a tabela de dados carregado.

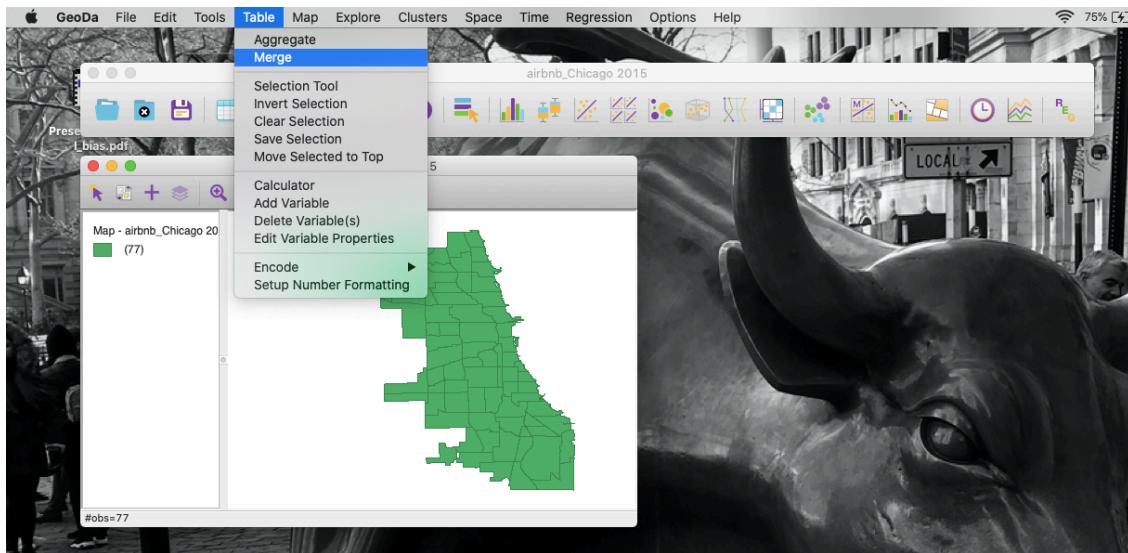


Figura 4: Iniciando merge de um conjunto externo de dados

Atente-se para o tipo de dados suportados pelo Geoda, como mostrado na Figura 5.

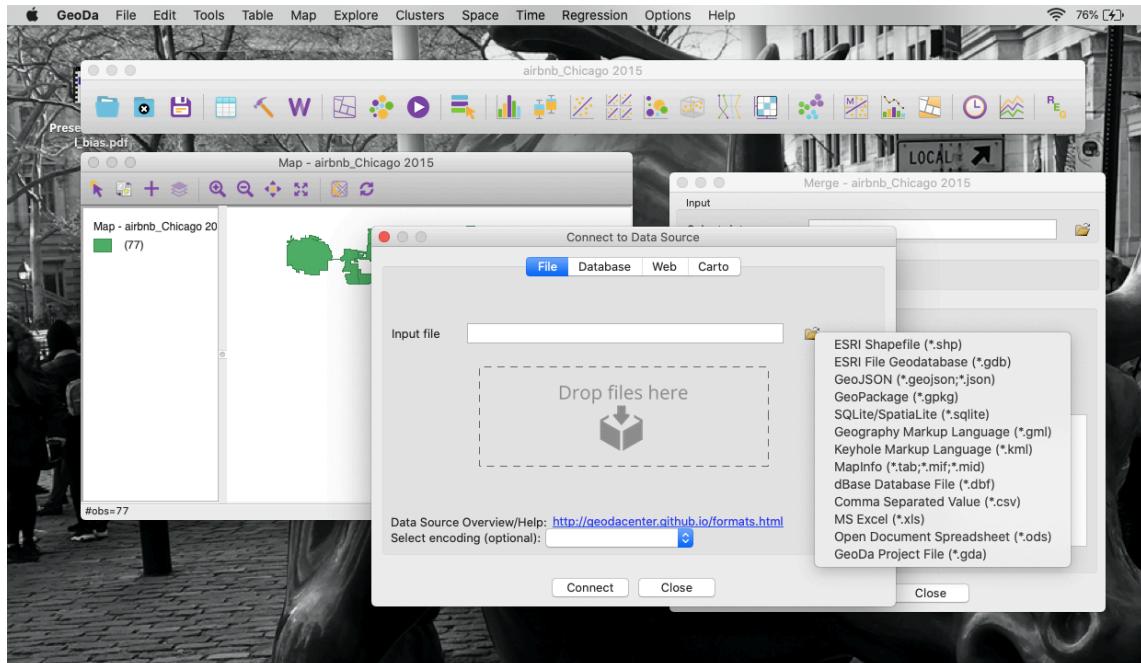


Figura 5: Tipos de dados suportados pelo Geoda.

Ao escolher os dados a serem unidos, certifique-se de que foram carregados corretamente.

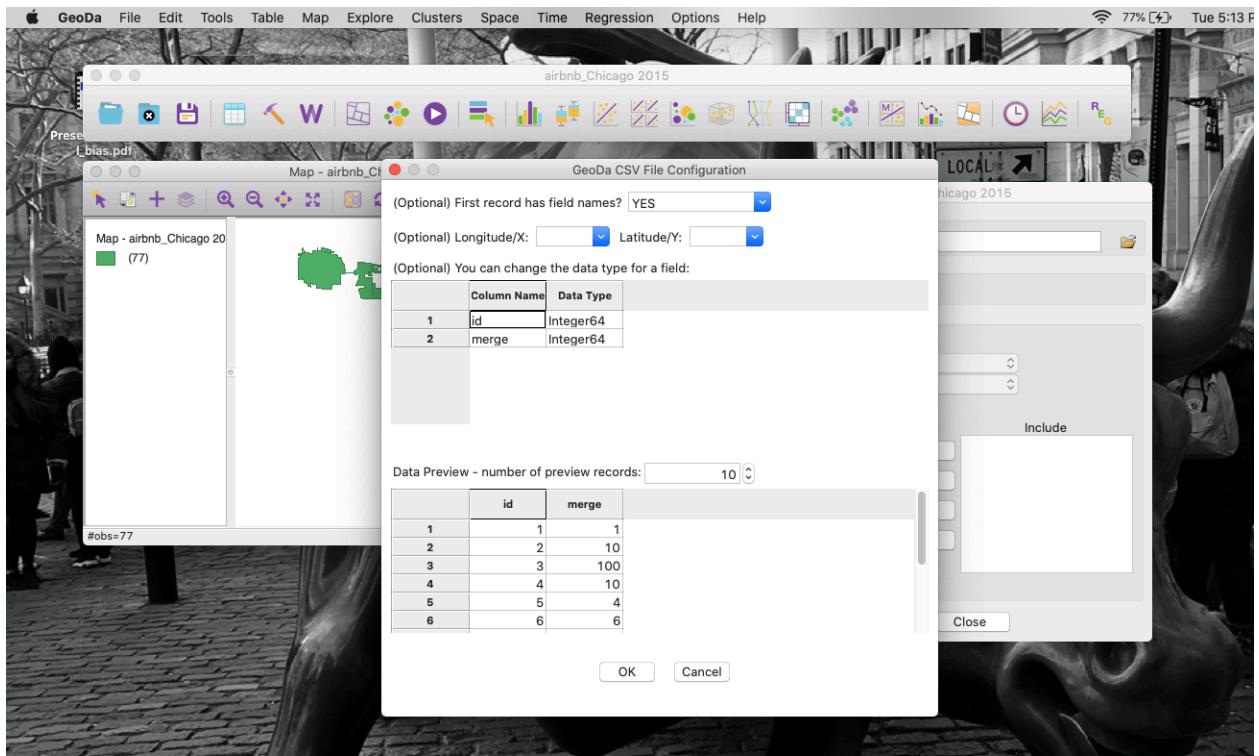


Figura 7: Carregando os dados corretamente

Para o sucesso do merge faz-se necessário que haja uma variável única que represente a mesma unidade espacial em ambos os conjuntos de dados. No exemplo, temos as variáveis *AREAID* e *id* sendo equivalentes. Por fim adicione as variáveis que deseja incluir no seu conjunto de dados na coluna a direita, conforme mostrado na figura 9.

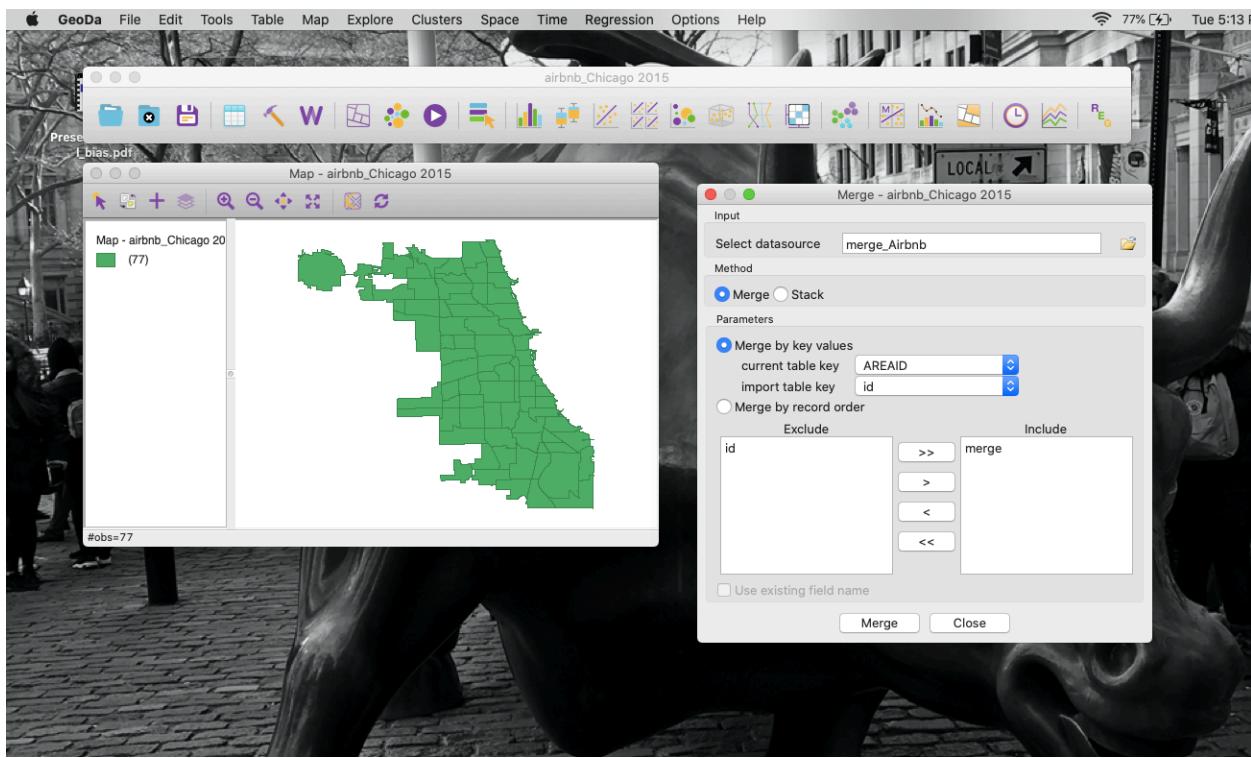


Figura 8: key variables

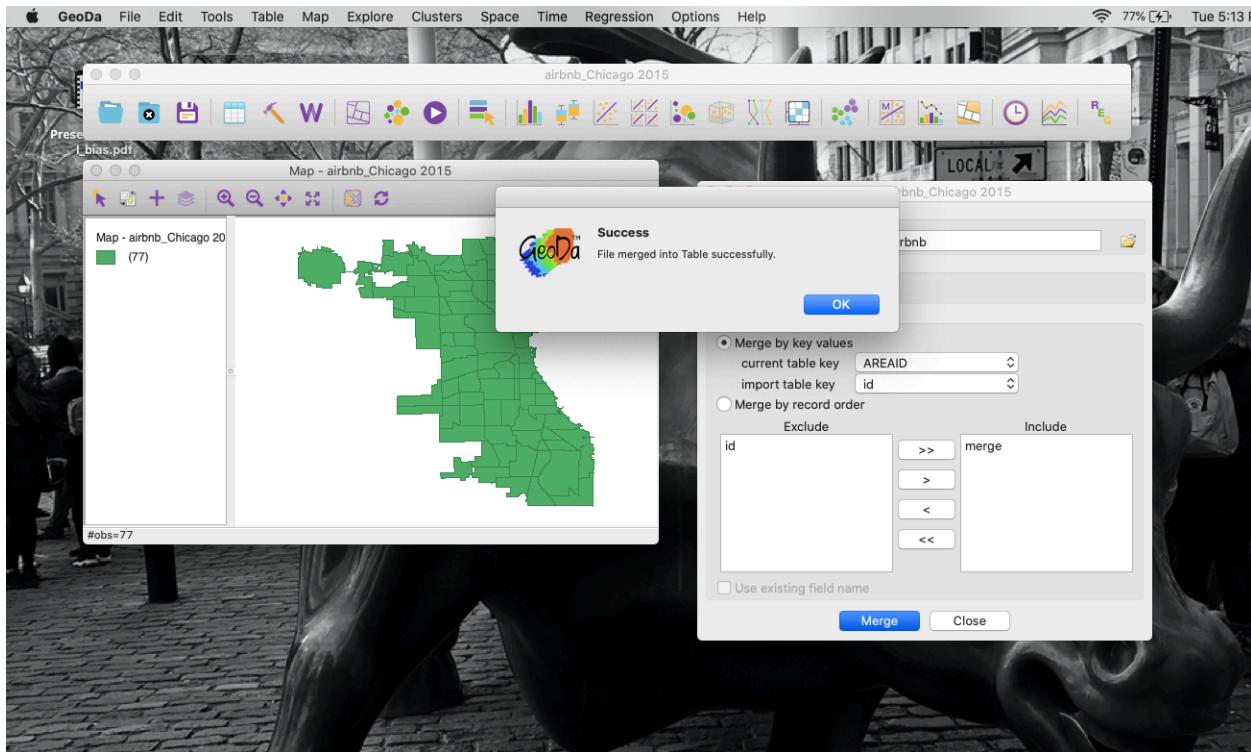


Figura 9: Merge finalizado com êxito.

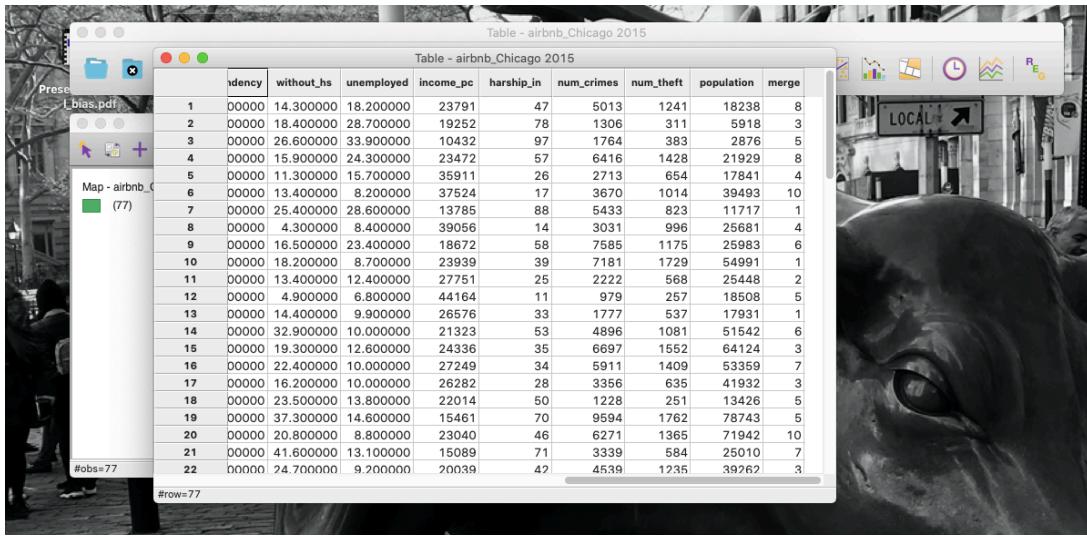


Figura 10: Conferindo se a variável foi incluída.

3. Plotting

Uma das principais vantagens do Geoda são as ferramentas de visualização. Na barra de opções em Map há diversas opções de mapas, tais como, mapas quantílicos, desvios-padrão, temáticos, entre outros.

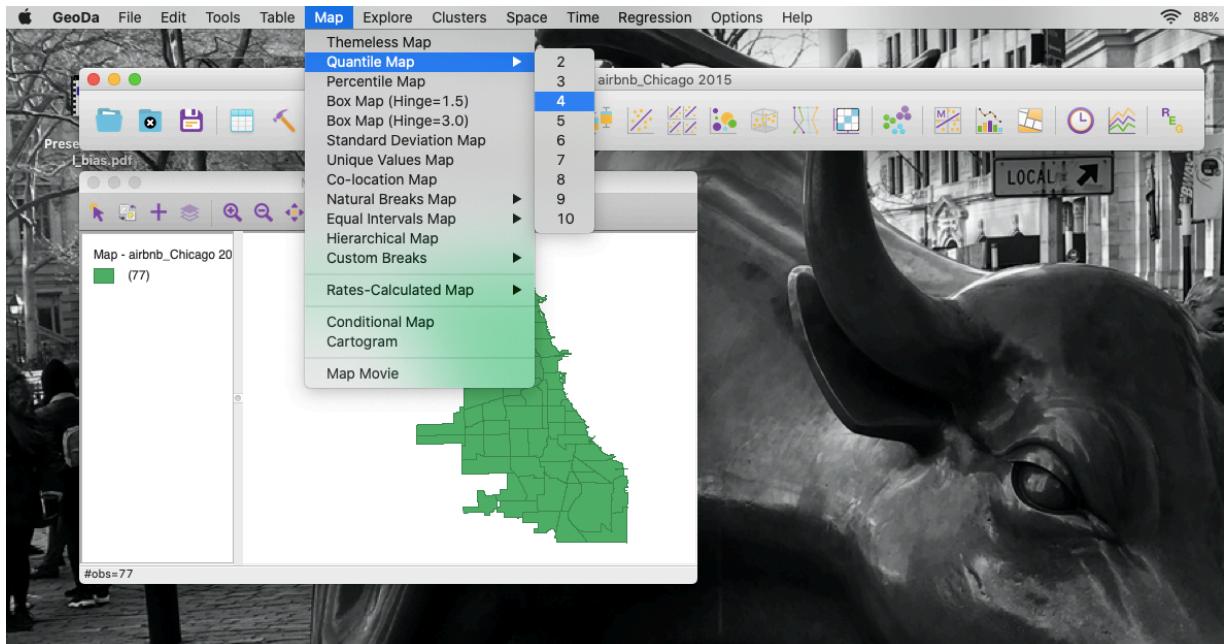


Figura 11: opções de mapas

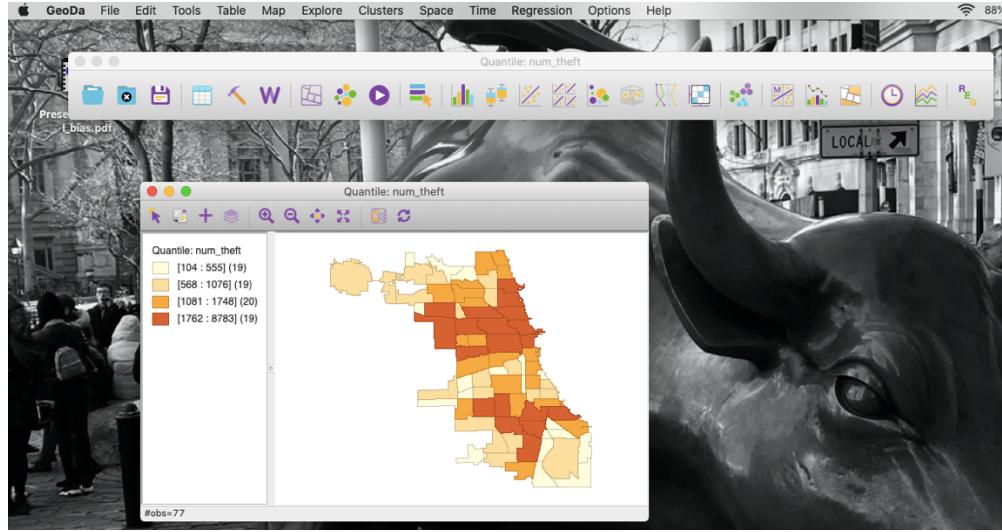


Figura 12: mapa quantílico para a variável de número de roubos

4. Diagramas de dispersão

Outra ferramenta de visualização muito útil é o diagrama de dispersão entre duas variáveis, no qual é possível analisarmos a relação linear entre as duas variáveis. No menu de opções selecione *Explore* e então *Scatter plot*.

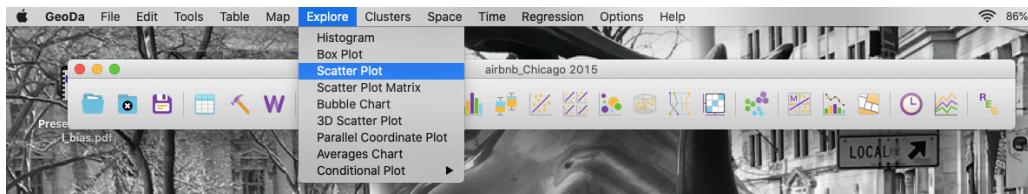


Figura 13: *Scatter plot*

Selecione as duas variáveis:

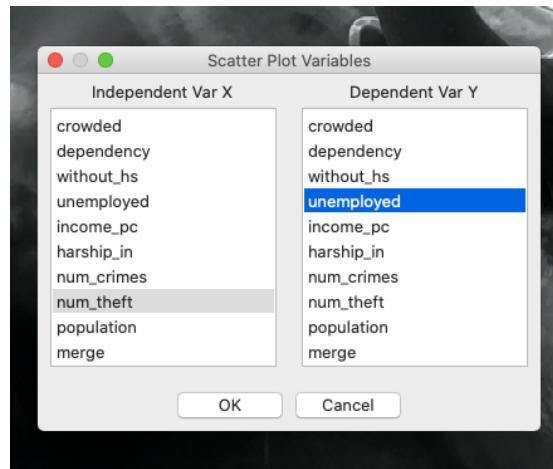


Figura 14: selecione as duas variáveis

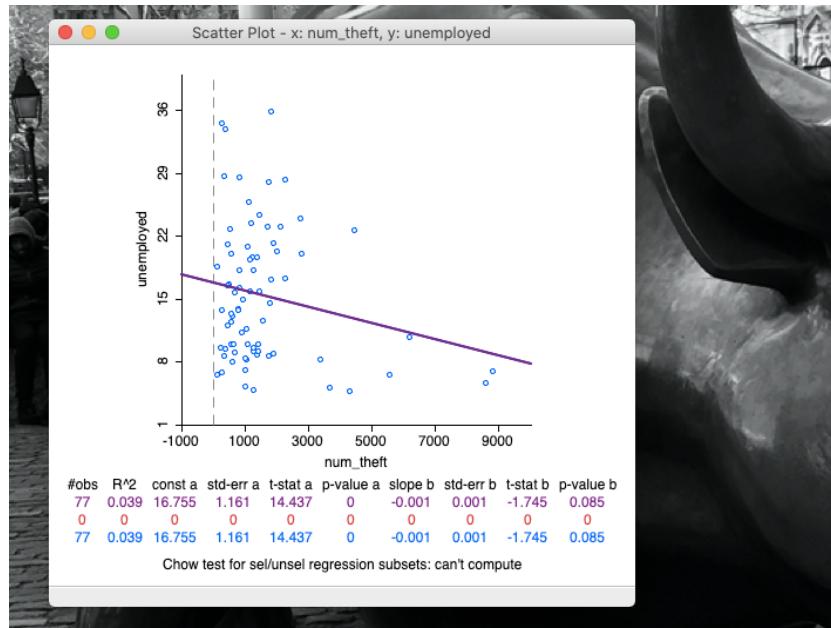


Figura 15: diagrama de dispersão entre número de crimes e desemprego

5. Matrizes espaciais

De maneira muito fácil e intuitiva, o Geoda fornece opções de criarmos matrizes de vizinhos mais próximos. No exemplo abaixo, uma matriz de contiguidade do tipo rainha de primeira ordem é criada. O primeiro passo consiste em clicar no símbolo *W (weights)* no software.

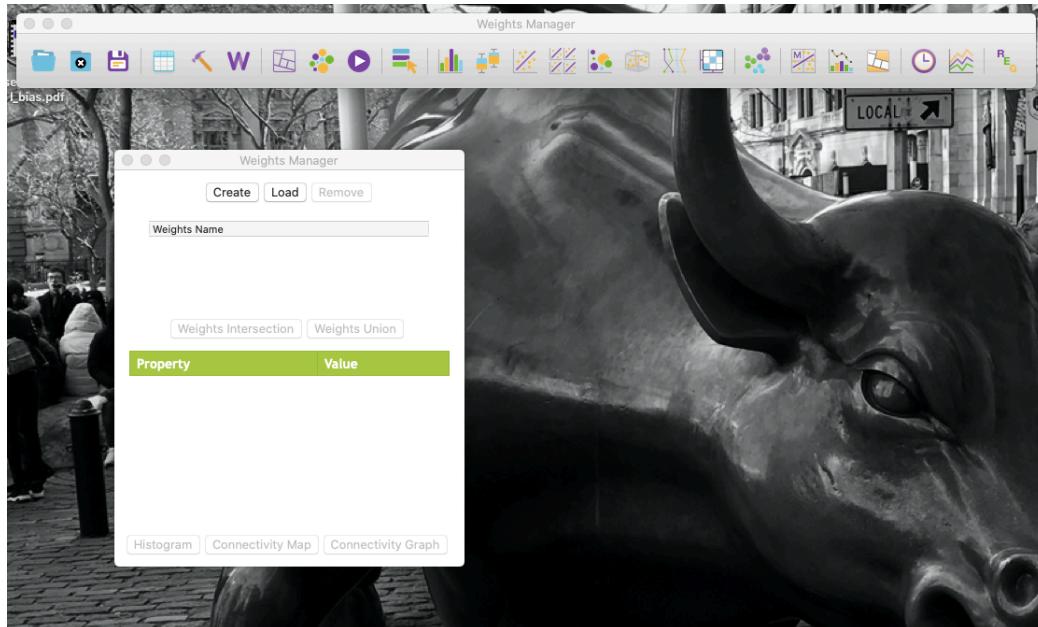


Figura 16: criando matrizes espaciais

Em seguida, certifique-se de escolher uma variável de valor único para a criação da matriz. Usualmente, utiliza-se a variável de identificação da unidade, como a variável *AREALD* utilizada no *merge*. Outros exemplos de variáveis chaves são os códigos do IBGE. Nesta etapa seleciona-se ainda o tipo de matriz a ser criada. Conforme dito anteriormente, está sendo criada uma matriz *queen* de ordem de contiguidade igual a 1.

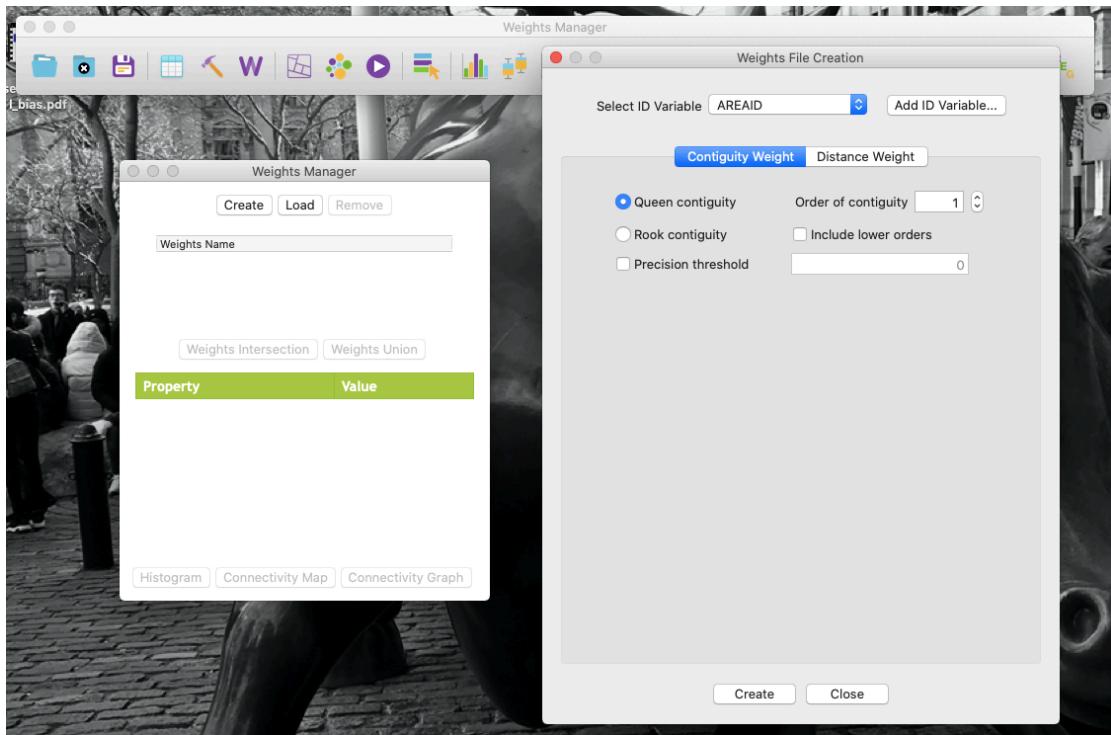


Figura 17: selecionando a variável chave, bem como o tipo de matriz.

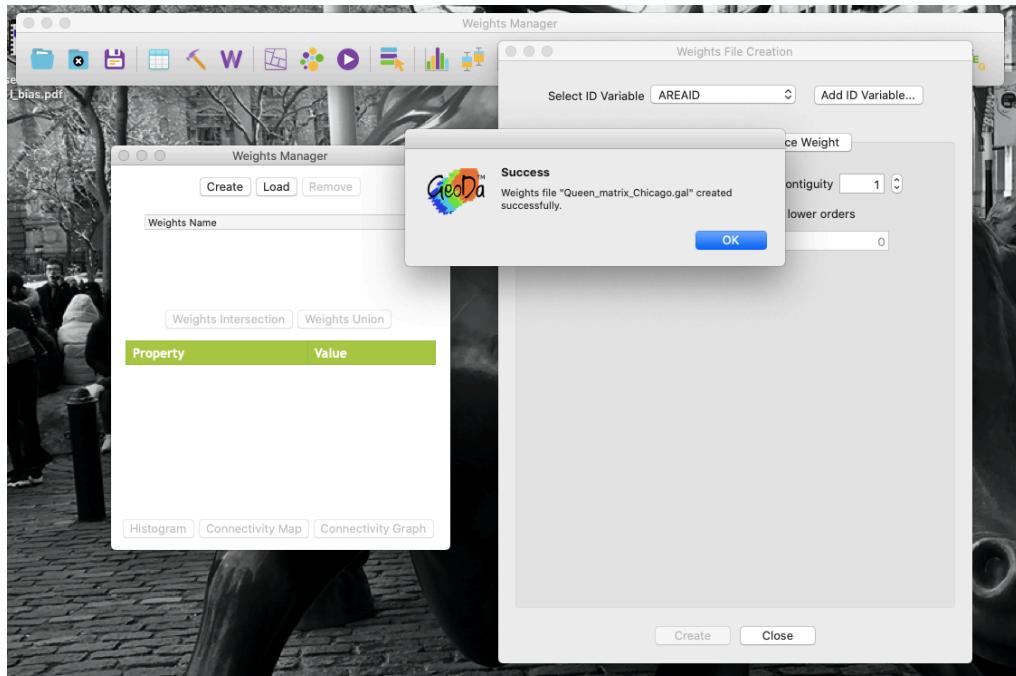


Figura 18: matriz criada com sucesso

Após finalizada a criação, um arquivo com a extensão gal foi gerado. Sendo possível ser carregado posteriormente, inclusive por outros meios, como R e Python. Estatísticas descritivas sobre a matriz podem ser observadas, bem como um mapa com as conectividades das vizinhas.

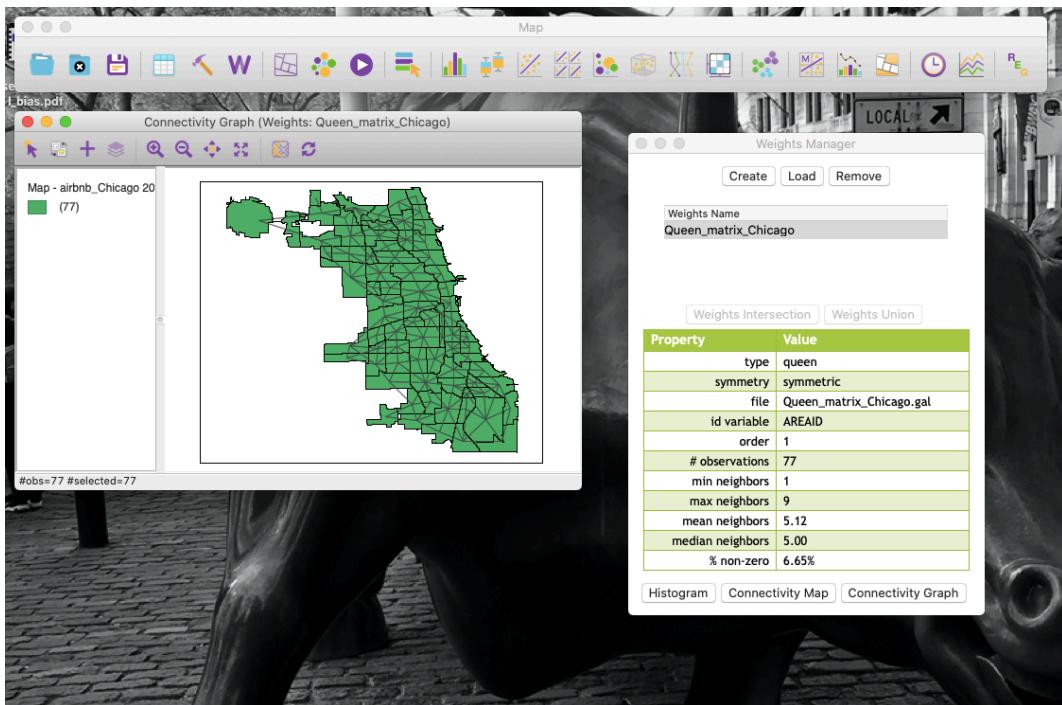


Figura 19: Conectividade da matriz

6. Calculator

Uma das maneiras fáceis de acrescentar novas variáveis no Geoda é utilizando o *calculator*.

As figuras abaixo mostraram como criar a variável de número de roubos defasada espacialmente utilizando a matriz *queen* criada. Para acessar o menu da Figura abaixo, clique com o botão direito do cursor no topo da tabela de dados.

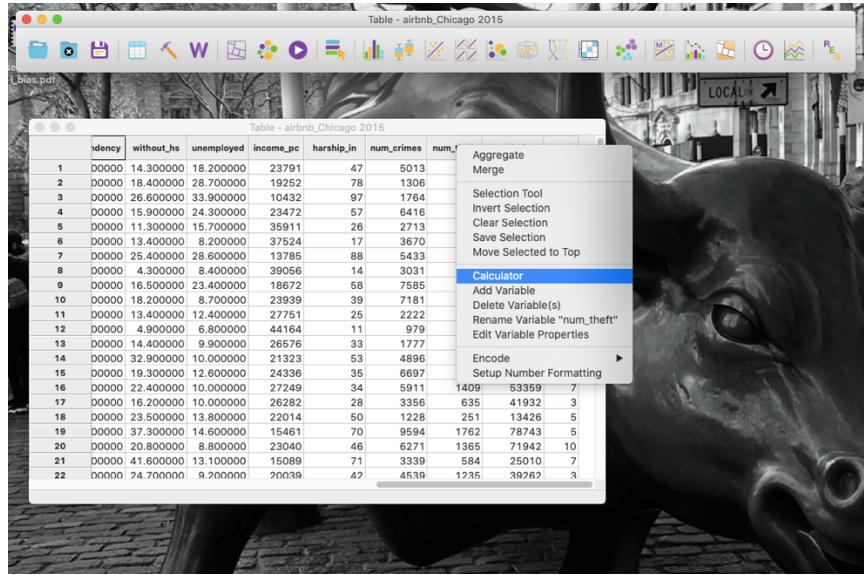


Figura 20: *calculator*

O primeiro passo é acrescentar uma nova variável a tabela de dados para receber os valores defasados espacialmente.

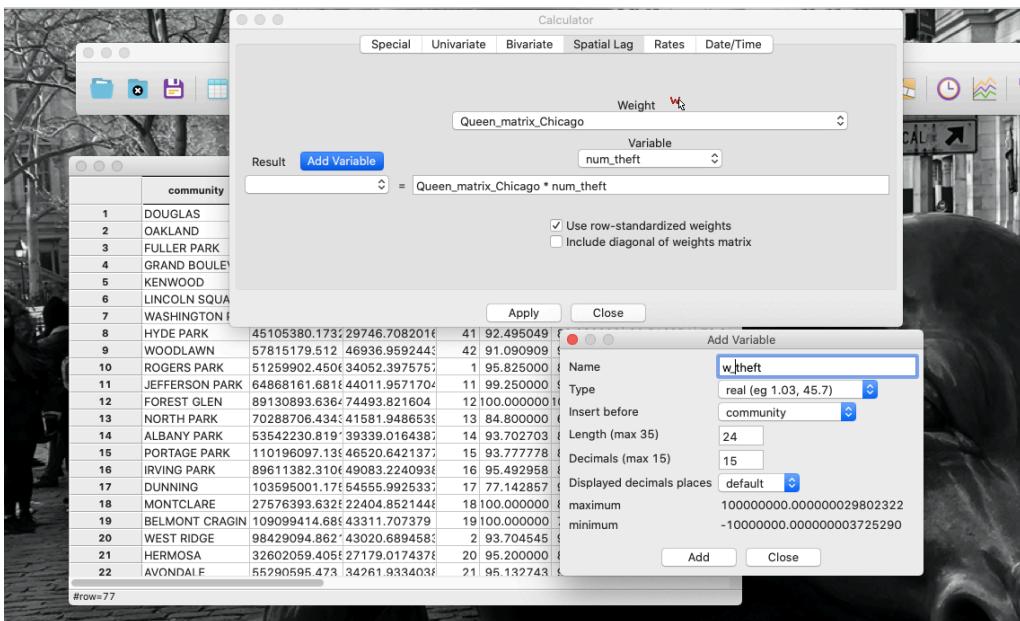


Figura 21: adicionando uma nova variável

Criada a variável escolhe-se a variável a ser defasada espacialmente, bem como a matriz de pesos espaciais a ser usada.

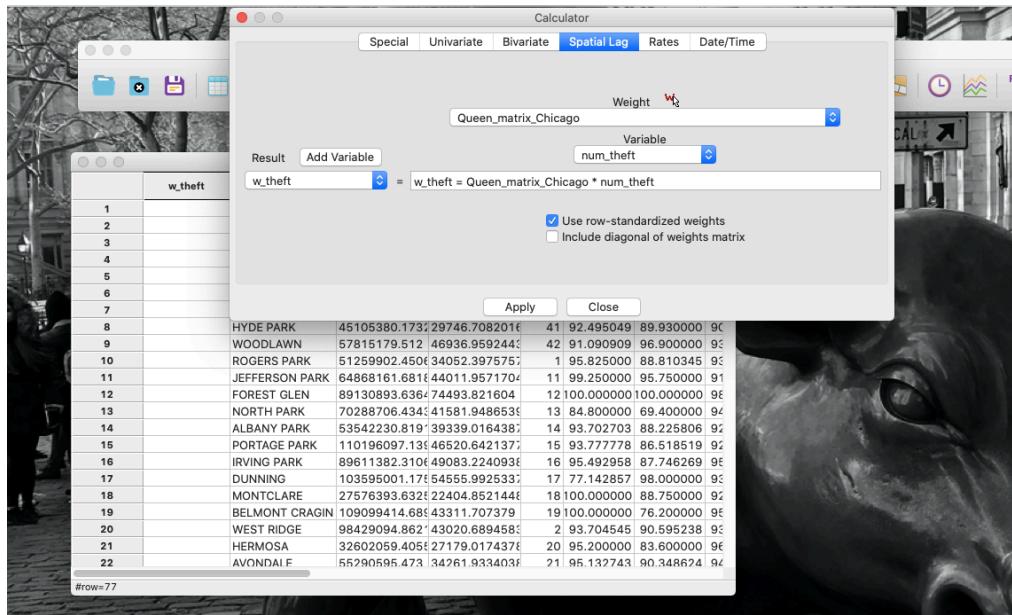


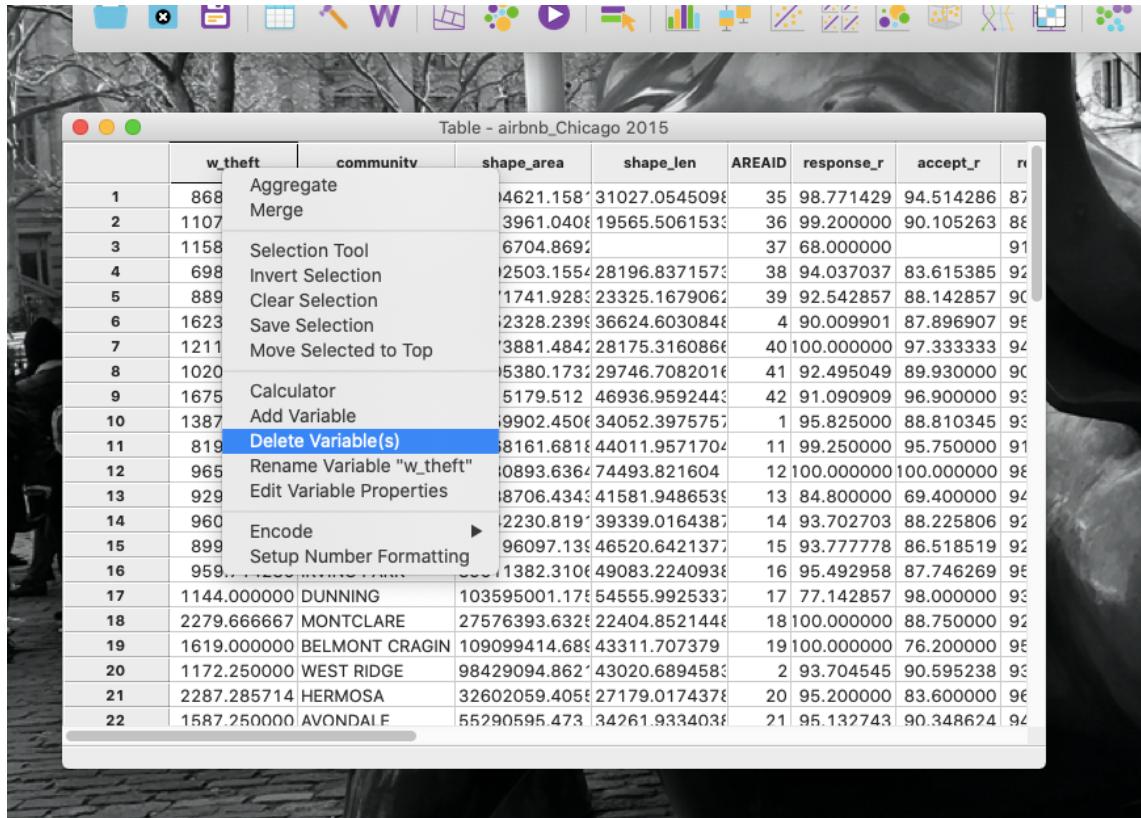
Figura 22: Defasando espacialmente a variável de número de roubos.

The screenshot shows a Jupyter Notebook cell with a table titled "Table - airbnb_Chicago 2015". The table has columns: w_theft, community, shape_area, shape_len, AREAID, response_r, accept_r, and r. The data consists of 77 rows, each representing a neighborhood in Chicago with its name, coordinates, and various statistical values. The last row of the table is "#row=77".

	w_theft	community	shape_area	shape_len	AREAID	response_r	accept_r	r
1	868.250000	DOUGLAS	46004621.158	31027.0545098	35	98.771429	94.514286	87
2	1107.666667	OAKLAND	169113961.0408	19565.5061533	36	99.200000	90.105263	88
3	1158.166667	FULLER PARK	199116704.8692		37	68.000000		91
4	698.857143	GRAND BOULEVAR	48492503.155	28196.8371573	38	94.037037	83.615385	92
5	889.500000	KENWOOD	29071741.9283	23325.1679062	39	92.542857	88.142857	90
6	1623.375000	LINCOLN SQUARE	71352328.2398	36624.6030848	4	90.009901	87.896907	95
7	1211.714286	WASHINGTON PAR	42373881.484	28175.3160866	40	100.000000	97.333333	94
8	1020.000000	HYDE PARK	45105380.1732	29746.7082016	41	92.495049	89.930000	90
9	1675.000000	WOODLAWN	57815179.512	46936.9592443	42	91.090909	96.900000	93
10	1387.000000	ROGERS PARK	51259902.4508	34052.397575	1	95.825000	88.810345	93
11	819.000000	JEFFERSON PARK	64868161.6818	44011.9571704	11	99.250000	95.750000	91
12	965.833333	FOREST GLEN	89130893.6364	74493.821604	12	100.000000	100.000000	98
13	929.250000	NORTH PARK	70288706.4343	41581.9486539	13	84.800000	69.400000	92
14	960.666667	ALBANY PARK	53542230.8197	39339.016438	14	93.702703	88.225806	92
15	899.428571	PORTAGE PARK	110196097.1371	46520.642137	15	93.777778	86.518519	92
16	959.714286	IRVING PARK	89611382.3106	49083.2240938	16	95.492958	87.746269	95
17	1144.000000	DUNNING	103595001.175	54555.9925337	17	77.142857	98.000000	93
18	2279.666667	MONTCLARE	27576393.6321	22404.8521448	18	100.000000	88.750000	92
19	1619.000000	BELMONT CRAGIN	109099414.688	43311.707379	19	100.000000	76.200000	95
20	1172.250000	WEST RIDGE	98429094.8621	43020.6894583	2	93.704545	90.595238	93
21	2287.285714	HERMOSA	32602059.405	27179.0174378	20	95.200000	83.600000	96
22	1587.250000	AVONDALE	55290595.473	34261.9334038	21	95.132743	90.348624	94

Figura 23: Verificando se a variável foi corretamente adicionada

É possível ainda deletar variáveis no conjunto de dados, um clique duplo com o mouse oferece a opção.



	w_theft	community	shape_area	shape_len	AREALD	response_r	accept_r	r
1	868	Aggregate	14621.158	31027.0545098	35	98.771429	94.514286	87
2	1107	Merge	3961.040	19565.5061533	36	99.200000	90.105263	86
3	1158	Selection Tool	6704.869		37	68.000000		91
4	698	Invert Selection	2503.155	28196.8371573	38	94.037037	83.615385	92
5	889	Clear Selection	1741.928	23325.1679062	39	92.542857	88.142857	90
6	1623	Save Selection	2328.239	36624.6030848	4	90.009901	87.896907	95
7	1211	Move Selected to Top	3881.484	28175.3160866	40	100.000000	97.333333	94
8	1020		5380.173	29746.7082016	41	92.495049	89.930000	90
9	1675	Calculator	5179.512	46936.9592443	42	91.090909	96.900000	93
10	1387	Add Variable	9902.450	34052.3975757	1	95.825000	88.810345	93
11	819	Delete Variable(s)	8161.681	44011.9571704	11	99.250000	95.750000	91
12	965	Rename Variable "w_theft"	0893.636	74493.821604	12	100.000000	100.000000	98
13	929	Edit Variable Properties	8706.434	341581.9486539	13	84.800000	69.400000	92
14	960	Encode	2230.819	39339.0164387	14	93.702703	88.225806	92
15	899	Setup Number Formatting	96097.139	46520.6421377	15	93.777778	86.518519	92
16	959		1382.310	49083.2240938	16	95.492958	87.746269	95
17	1144.000000	DUNNING	103595001.175	54555.9925337	17	77.142857	98.000000	93
18	2279.666667	MONTCLARE	27576393.632	22404.8521448	18	100.000000	88.750000	92
19	1619.000000	BELMONT CRAGIN	109099414.689	43311.707379	19	100.000000	76.200000	95
20	1172.250000	WEST RIDGE	98429094.862	43020.6894583	2	93.704545	90.595238	93
21	2287.285714	HERMOSA	32602059.405	27179.0174378	20	95.200000	83.600000	96
22	1587.250000	AVONDALE	55290595.473	34261.9334038	21	95.132743	90.348624	94

Figura 24: deletando uma variável

7. I de Moran

7.1 I de Moran global univariado

O primeiro passo na análise de dados espaciais é testar a hipótese de aleatoriedade espacial. Para dados areais, a Estatística de I de Moran é amplamente utilizada. O coeficiente pode ser pensado como a autocorrelação espacial entre uma região e suas regiões adjacentes.

Um valor positivo do coeficiente indica uma relação de similaridade entre as unidades, enquanto um valor negativo evidencia uma dissimilaridade.

O Geoda oferece uma maneira simples para o cálculo do índice. Clique em *Space* na barra de opções e a opção, em seguida em *Univariate Moran's I*.

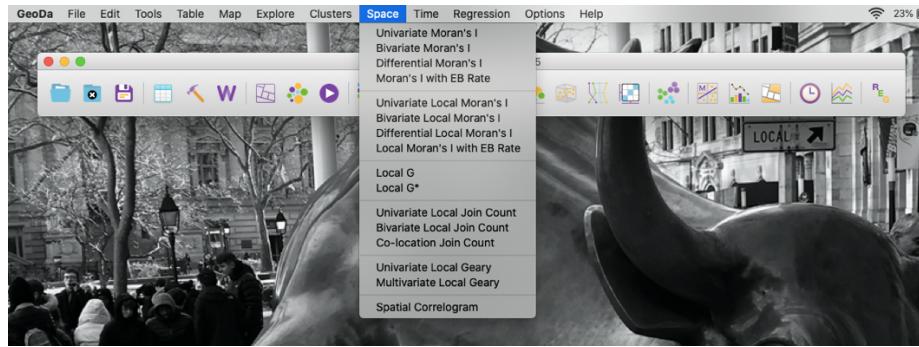


Figura 25: I de Moran univariado

Em seguida escolha a variável que deseja calcular o I de Moran, bem como a matriz de pesos espaciais para o cálculo.

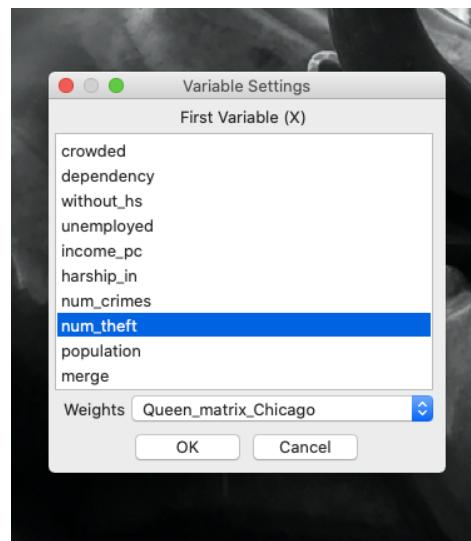


Figura 26: Escolha da variável e matriz para o cálculo do I de Moran



Figura 27: Gráfico de dispersão do I de Moran

Pelo resultado, percebe-se que a variável de número de roubos apresenta um I de Moran positivo, o que indica uma associação de similaridade entre as unidades espaciais. Em outras palavras, regiões com um número alto de crimes contra a propriedade está cercada por regiões com alto número de roubos. Da mesma forma, regiões com baixo número de roubos está cercada por regiões com baixo número de roubos.

Há algumas formas de se avaliar a significância estatística do I de Moran, tais como assumira normalidade dos dados e extrair informações sobre a mesma utilizando a função de distribuição acumulada da mesma, ou *bootstrap*.

No entanto, a mais usual é utilizar a permutação aleatória. Todos os valores observados para uma variável são permutados um número n de vezes, e para cada permutação calcula-se o I de Moran. Desta forma, obtém-se uma distribuição dos valores permutados aleatoriamente, permitindo a computação inferencial sobre os dados observados. O Geoda oferece a opção de permutação ao clicar com o botão direito sob o gráfico de dispersão, em seguida clique em *randomization* e escolha o número de permutações. Recomenda-se que quanto maior o tamanho amostral analisado, maior o número de permutações.

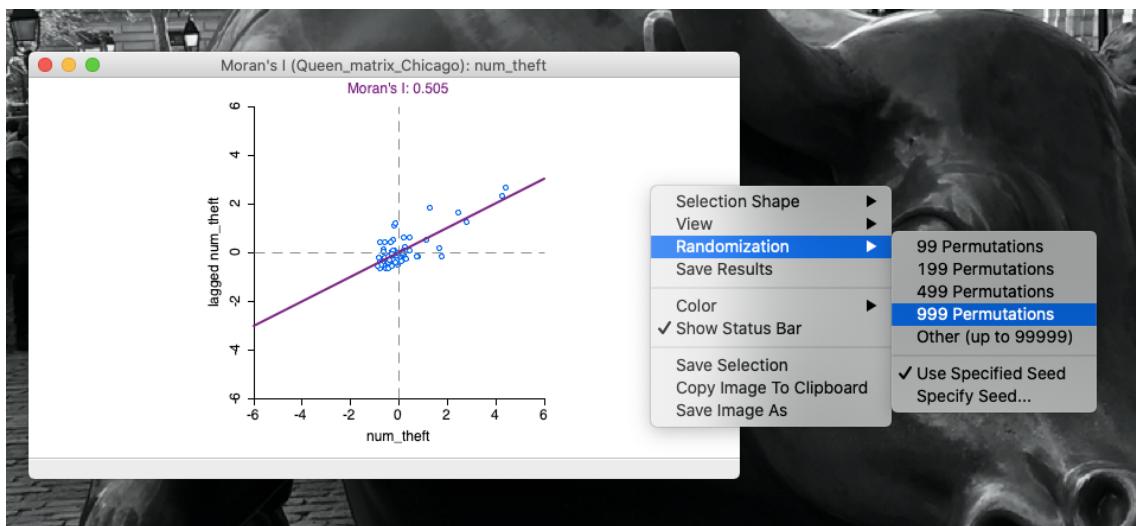


Figura 28: permutação

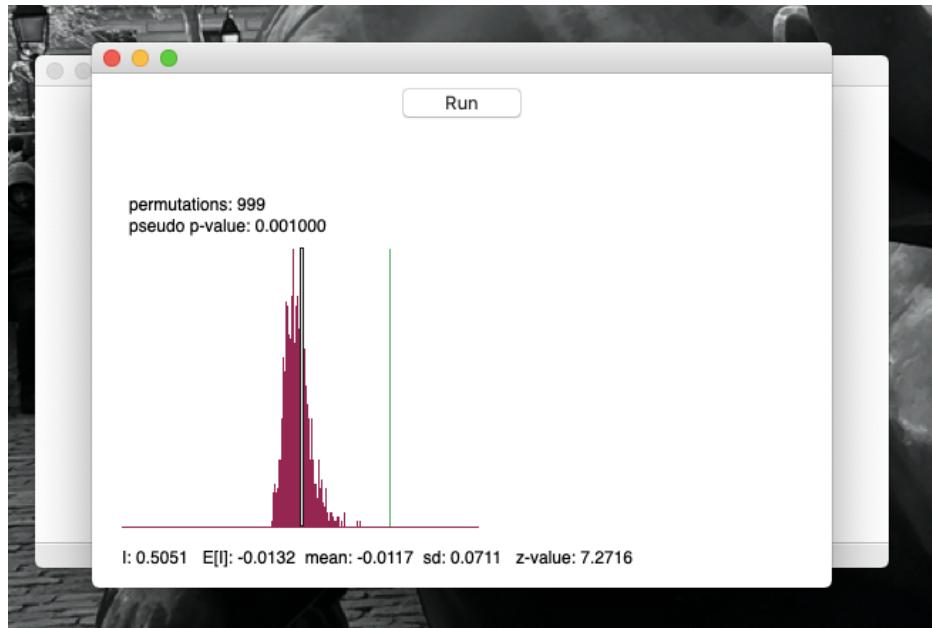


Figura 29: significância estatística do I de Moran univariado

Pela figura 29, rejeitamos a hipótese nula de aleatoriedade espacial a níveis convencionais de significância estatística. Concluindo que ao modelar o número de roubos em Chicago para o ano de 2015 há evidências de que efeitos espaciais devem ser considerados na modelagem econométrica.

7.2 I de Moran global bivariado

A ideia de um indicador espacial bivariado é analisar se há alguma relação com as observações de uma dada variável em uma unidade espacial com outra variável nas regiões adjacentes. Utilizando o I de Moran global bivariado para a mesma, tem-se na barra de opções conforme a Figura 30 abaixo.

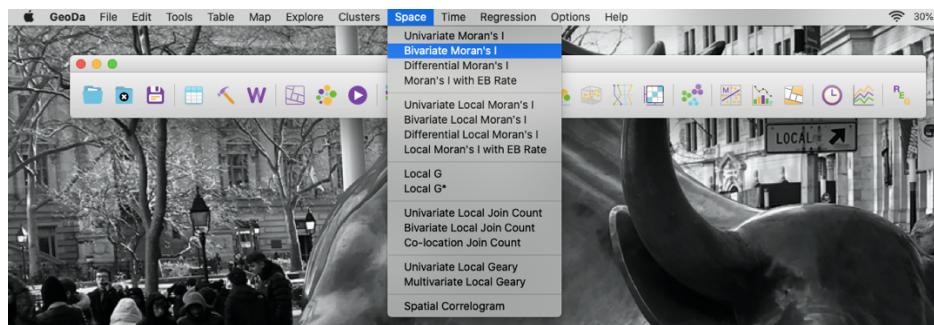


Figura 30: I de Moran bivariado

Seguindo o que já foi feito anteriormente, abaixo encontra-se o I de Moran bivariado entre o número de roubos e a taxa de desemprego. Note que a variável de taxa de desemprego encontra-se defasada espacialmente utilizando a matriz de pesos espaciais do tipo rainha criada anteriormente. Ou seja, a análise consiste em avaliar como o número de roubos em uma dada região está relacionado com a taxa de desemprego das regiões vizinhas.

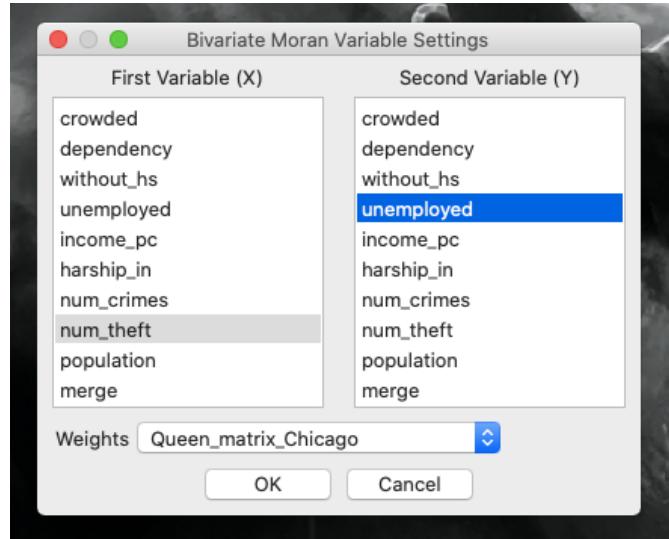


Figura 31: selecionando as variáveis e matriz de pesos para o cálculo do I de Moran global bivariado

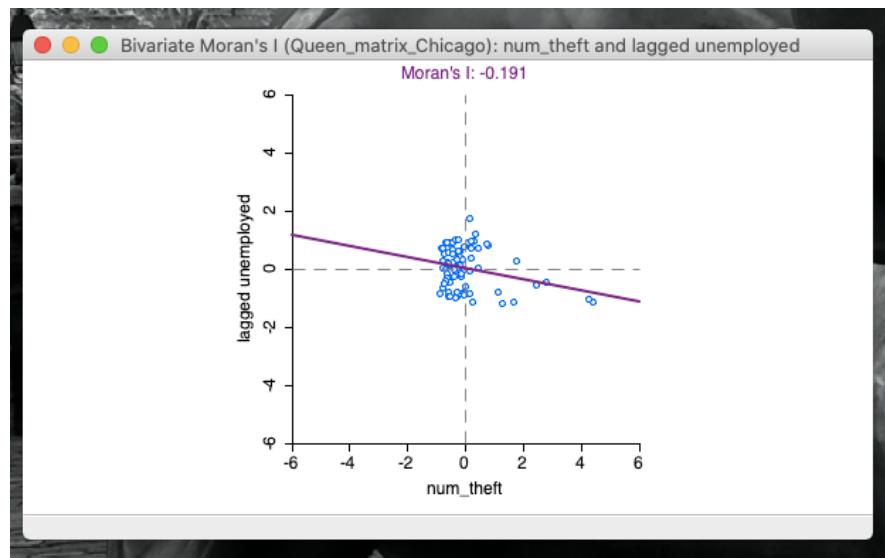


Figura 32: I de Moran global bivariado entre crime e desemprego, Chicago 2015

Note que há uma relação negativa, ou seja, as regiões com alto número de crimes contra a propriedade tendem a estar rodeadas por regiões com taxas baixas de desemprego, bem como

regiões com baixo número de roubos tendem a estar rodeados por regiões com alto número de desempregados.

Utilizando a permutação aleatória da mesma forma que para o I de Moran global univariado, encontra-se evidências estatisticamente significativas a 0.01% de rejeitarmos a hipótese nula de aleatoriedade espacial.

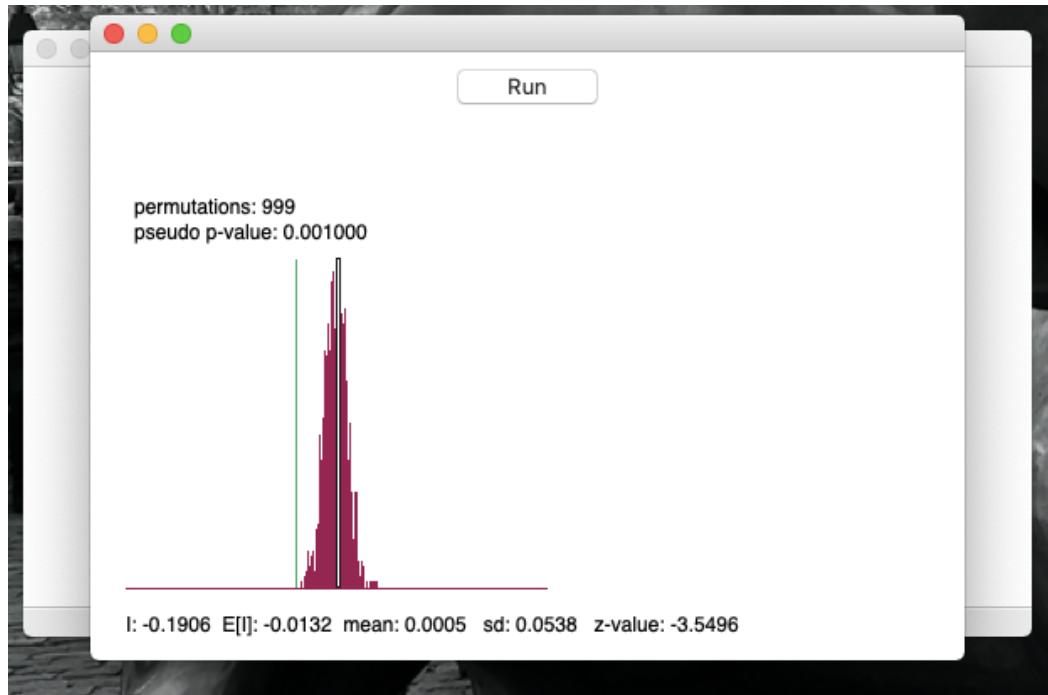


Figura 33: permutação aleatória