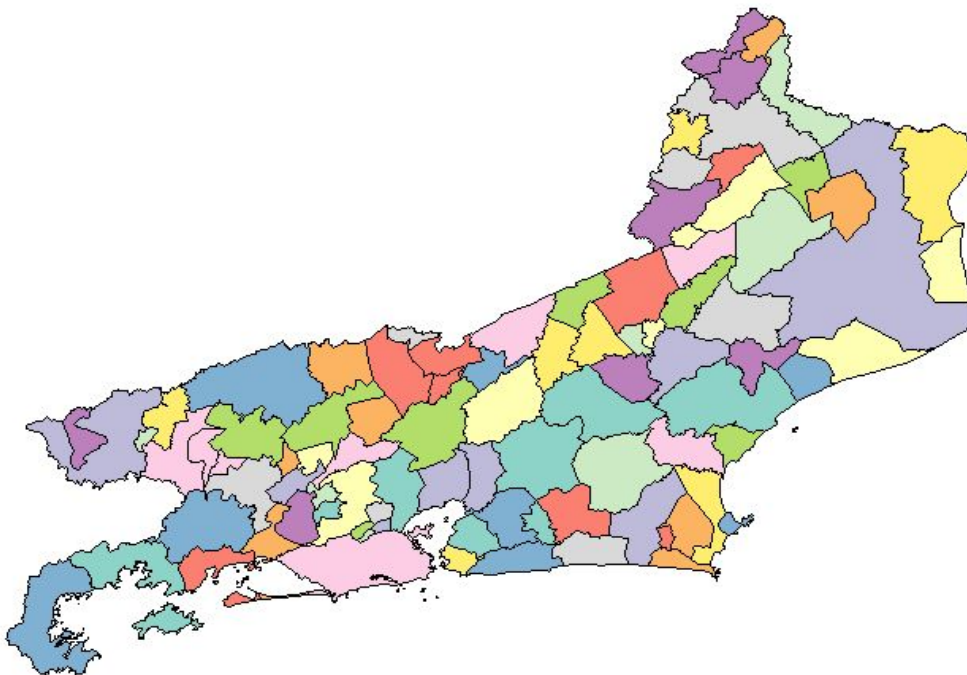


Trabalho: Aplicações em Estatística Espacial**Nome: Hugo Muniz Albuquerque**

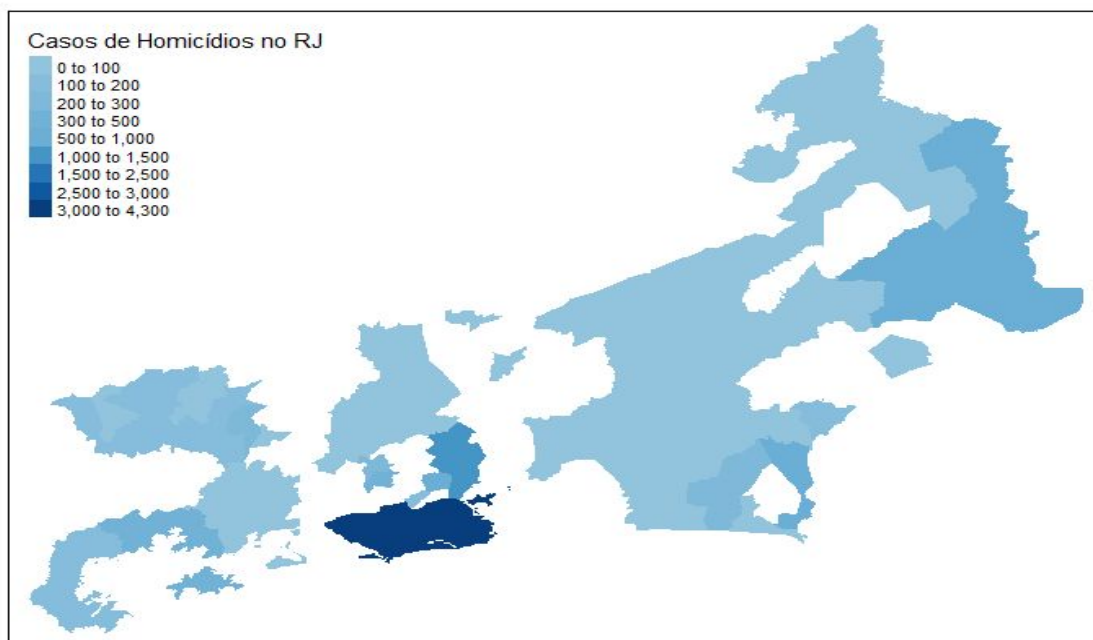
Neste trabalho vamos analisar dados de área para casos de homicídios nas cidades do estado do rio de janeiro. Abaixo podemos ver o mapa do espaço de interesse:

nome



Após esta apresentação do espaço, conferi se existia valores faltantes na base de dados e removi os dados faltantes encontrados. Em seguida, manipulei o extrato da base de dados, para obter a variável resposta homic em função de cada cidade e as variáveis explicativas proporção de homens e mulheres, proporção de pessoas com idade entre 31 e 40 anos e o IDH por cidade. Assim, foi possível analisar se existia uma dependência espacial.

Para analisar a distribuição espacial dos casos de homicídios entre as cidades do Rio de Janeiro, foi feito um mapa de dados de área que pode ser visto abaixo:



Em seguida, foi analisado a correlação entre as variáveis da base de dados.



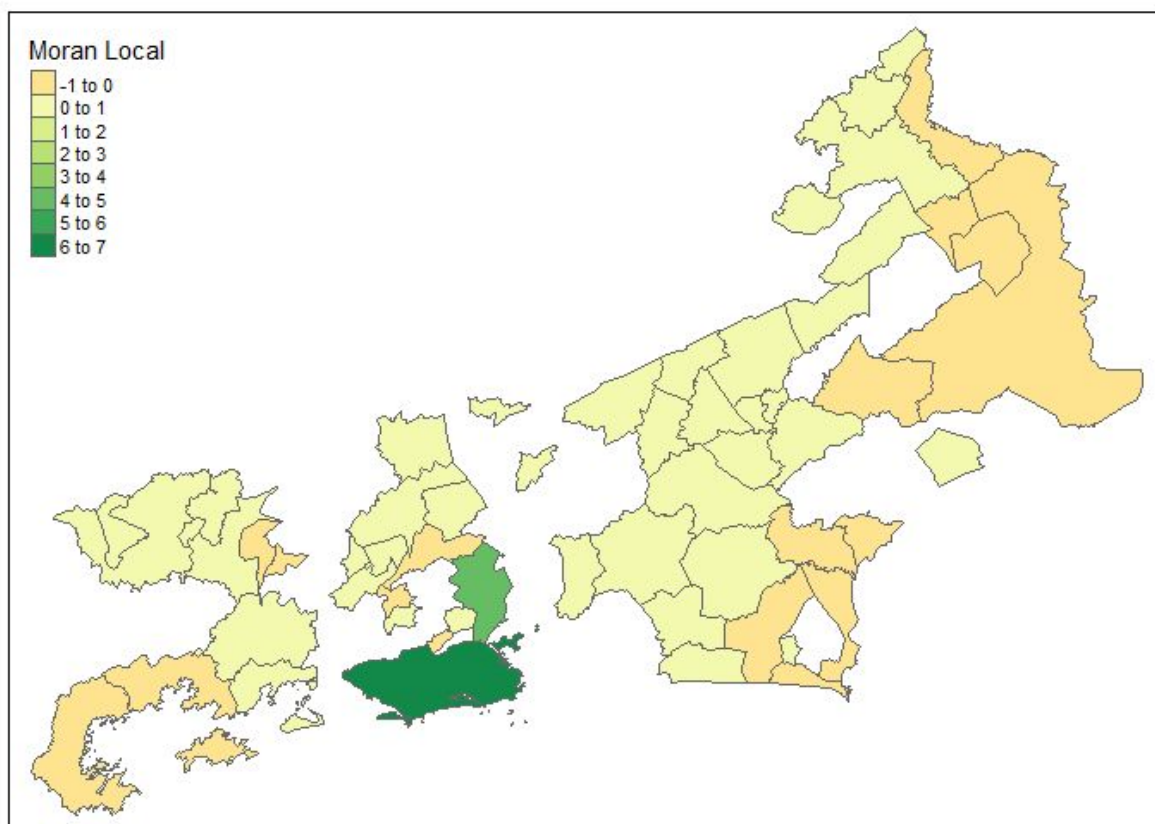
Neste gráfico de correlação do R podemos observar que a variável resposta homic possui uma correlação razoável em relação às variáveis prop_m, prop_h,

prop_idade_31a40 e IDH. Já entre as variáveis explicativas, nota-se uma correlação mais forte entre IDH e prop_m e prop_h o que pode prejudicar o modelo.

Para analisar a autocorrelação global e local foi criado a matrix de vizinhança com o critério Queen e depois com pesos padronizados por linhas, pesos binários e com 3 vizinhos para cada região e peso padronizado por linha. Após isso, foi aplicado o teste Moral Global nas seguintes matrizes de vizinhança com critério Queen: (1) com pesos padronizados por linhas, (2) com pesos binários e (3) com 3 vizinhos e padronizados por linhas. O resultado foi o seguinte respectivamente:

A primeira (1) matriz com pesos padronizados por linhas obteve um índice positivo de 0.22 e um pvalor de $1.172e-05$, o que é menor que o nível de significancia de 5%, portanto pode-se rejeitar a hipótese nula de que existe aglomerações de regiões vizinhas que possuem comportamento semelhante, ou seja, existe um padrão espacial para o número de casos de homicídios no Estado do Rio de Janeiro. A segunda (2) matriz com pesos binários obteve um índice positivo de 0.16 e um pvalor de 0.0002187, existe um padrão espacial para o número de casos de homicídios no Estado do Rio de Janeiro. A terceira (3) matriz com 3 vizinhos e padronizados por linhas obteve um índice positivo de 0.09 e um pvalor de 0.01883, existe um padrão espacial para o número de casos de homicídios no Estado do Rio de Janeiro. Portanto é possível modelar com base nessas 3 matrizes.

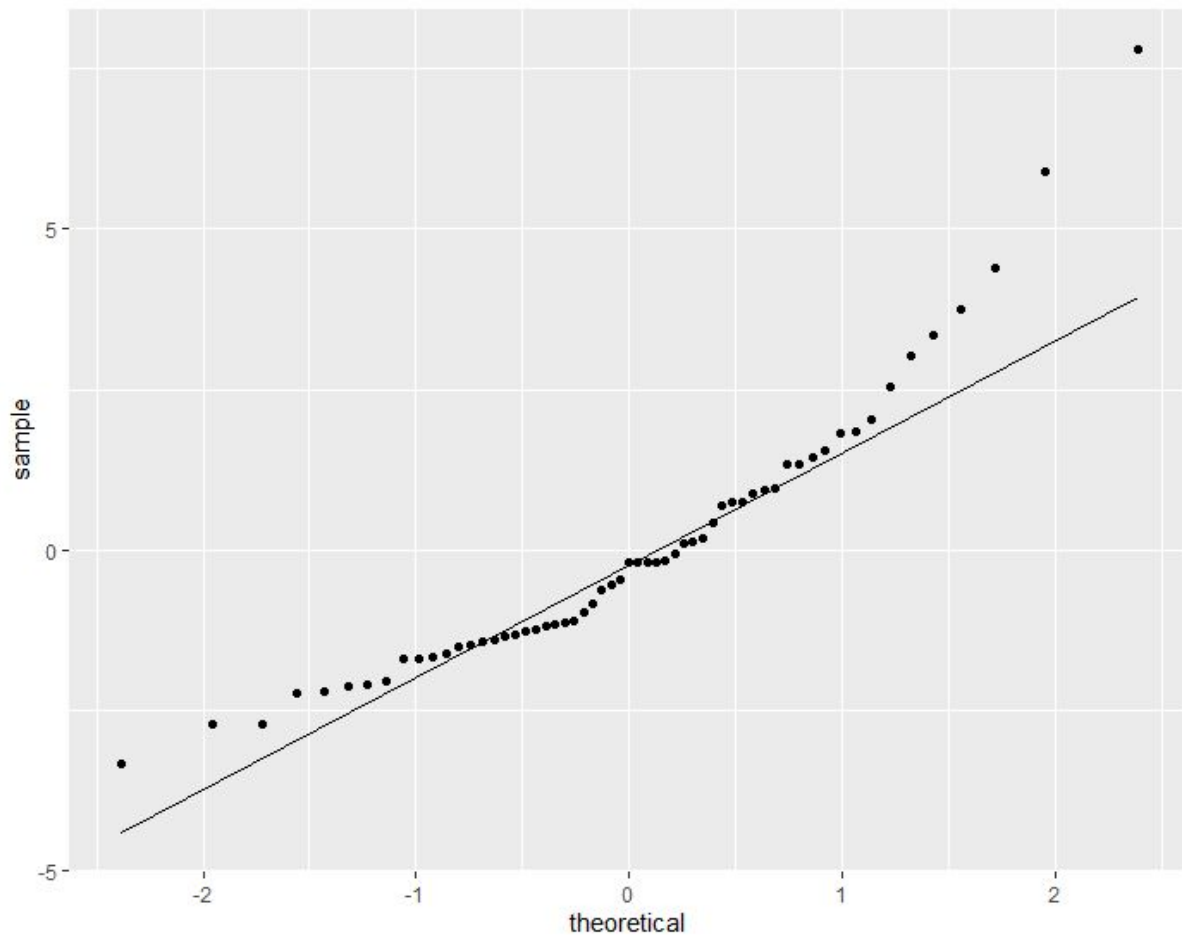
Além da análise do teste de hipótese de Moran, foi criado um mapa dos índices de moran local de cada cidade para analisar onde existem maiores concentrações de casos de homicídios.



No mapa acima pode-se observar que na capital do Estado do Rio de Janeiro existe uma maior aglomeração de vizinhos com comportamento semelhantes com a capital.

Por fim, foram ajustados alguns modelos, onde o melhor modelo encontrado foi o modelo SAR com matriz com critério Queen e pesos padronizados por linha e com as seguintes variáveis: *prop_m*, *prop_idade_31a40* e IDH. Além disso, foi aplicado uma transformação na variável resposta *homic*, elevando ela a um terço ($\frac{1}{3}$) para que fosse possível obter um pvalor significativo menor que 5%. O modelo obteve um pvalor de 0,0441, e os coeficientes indicam, respectivamente: (1) *prop_m* indica que a média de homicídios aumentam 0.8 quando ocorre um aumento de 1% na população de mulheres na cidade, (2) *prop_idade_31a40* indica que a média de homicídios aumenta 0.80 quando a proporção de pessoas com idade entre 31 a 40 anos aumenta 1%, e (3) IDH indica que quanto maior o IDH maior o número de casos de homicídios. Este modelo obteve a menor AIC e o menor erro quando comparado com o modelo de regressão linear múltipla, e os modelos SAR e CAR. A

variável IDH obteve um pvalor que indica não significância da variável, porém ao retirar a variável IDH o modelo passava a apresentar um pvalor de 0.12 o que tornava o modelo estatisticamente não significativo. Por último, vamos analisar a normalidade e a independência dos resíduos e plotar os resíduos dentro do shapefile do Rio de Janeiro.



No gráfico acima pode-se observar que os pontos se aproximam da linha da normal, o que indica normalidade dos resíduos. Abaixo veremos o texto de Moran 1 para os resíduos a fim de identificar se existe ou não independência dos resíduos do modelo:

Moran I test under randomisation

data: residuals(ajusteSAR2w)

weights: W.Queen.pesoW n reduced by no-neighbour observations

Moran I statistic standard deviate = -0.0075741, p-value = 0.503

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
-0.01899905	-0.01818182	0.01164226

Este teste nos mostra que a hipótese nula de independência dos resíduos não é rejeitada, pois o pvalor obtido foi superior a 5%, logo o modelo é válido. Então conclui-se que o melhor modelo é o SAR com peso padronizado por linhas e a maior concentração de casos de homicídios se encontra na capital e nas proximidades.