
Análise de Regressão Espacial

Aurelienne Jorge, Cristiane Brandão, Leonardo Rossatto, Nayla Martins,
Raíssa Teixeira, Renata Quevedo, Tatiana Uehara, Thiago Bazzan

Disciplina: Análise Espacial de Dados Geográficos (SER-301)
Professores Responsáveis: Dr. Eduardo G. Camargo e Dr. Antônio Miguel Vieira Monteiro

Regressão Espacial

Ferramenta estatística que utiliza a relação entre duas ou mais variáveis tal que uma variável possa ser explicada pela(s) outra(s).

Objetivos:

- Determinar relação entre duas ou mais variáveis;
- Estimar uma função que defina a variável dependente (Y);
- Possibilitar a previsão de valores futuros para a variável dependente.

Modelo Spatial Lag

$$Y = \rho WY + X\beta + \varepsilon$$

ρ = coeficiente espacial autoregressivo - medida de correlação espacial ($\rho = 0$, se autocorrelação é nula - hipótese nula)

W = matriz de proximidade espacial

WY expressa a dependência espacial em Y

A variável Y_i é afetada pelos valores da variável resposta nas áreas vizinhas a i .

Dependência em relação às **variáveis originais** nas áreas vizinhas – lag dependence /missing error

Modelo Spatial Error

$$Y = X\beta + \varepsilon$$
$$\varepsilon = \rho W\varepsilon + \xi$$

$W\varepsilon$ = erro com efeitos espaciais

ρ = medida de correlação espacial

ξ = componente do erro com variância constante e não correlacionada.

Dependência em relação aos **resíduos** nas áreas vizinhas - error dependence / missing lag

GWR - Geographically Weighted Regression

$$Y(i) = \beta(i)X + \varepsilon$$

$Y(i)$: variável que representa o processo no ponto i .

$\beta(i)$:parâmetros estimados no ponto i .

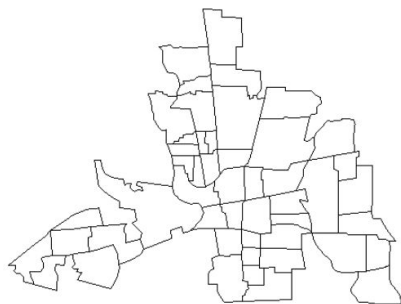
Ajusta o modelo de regressão a cada ponto observado, ponderando todas as demais observações como função da distância a este ponto.

Modelos Locais e Globais.

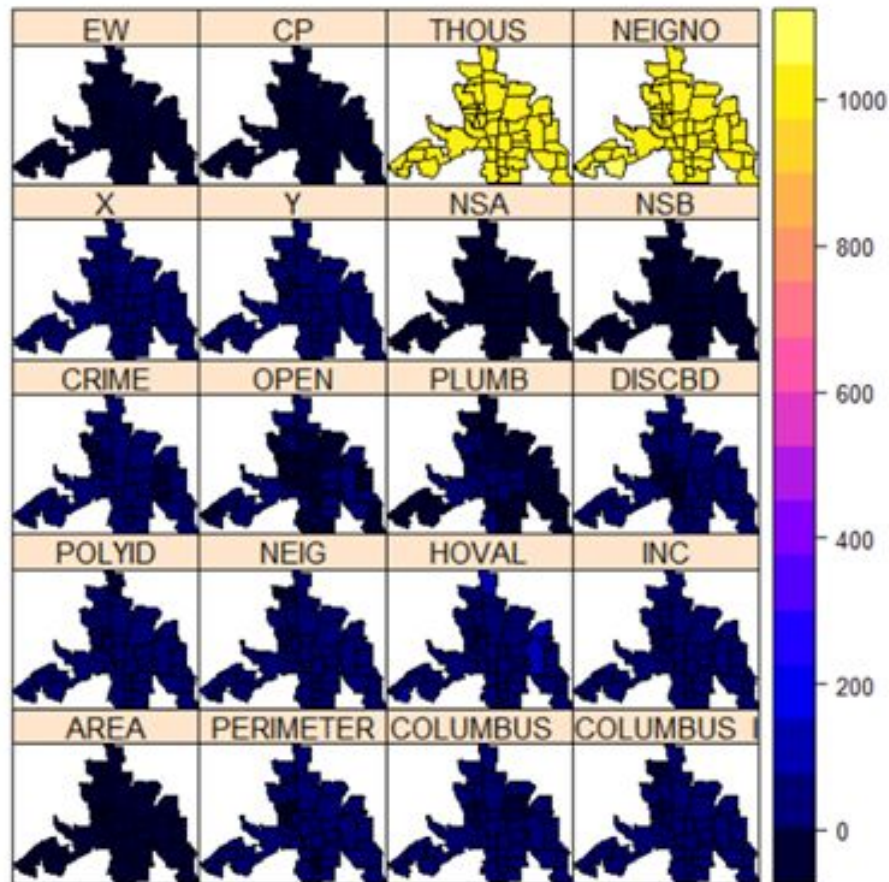
Pontos mais próximos ao ponto central possuem maior peso.

Resultados

Análise exploratória dos dados

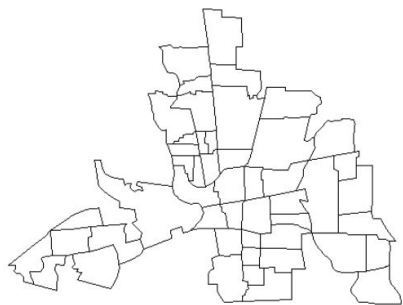


Columbus

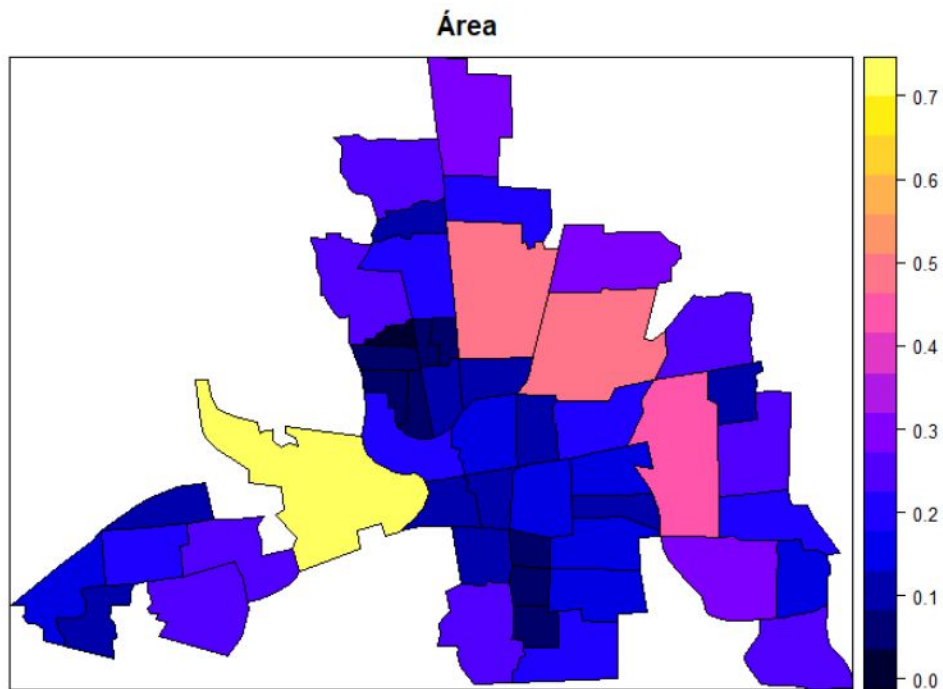


Resultados

Análise exploratória dos dados

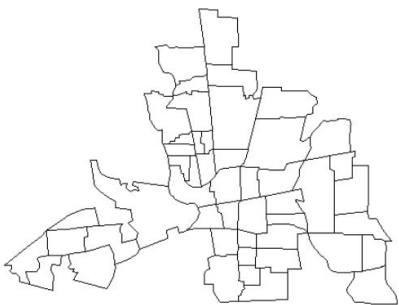


Columbus



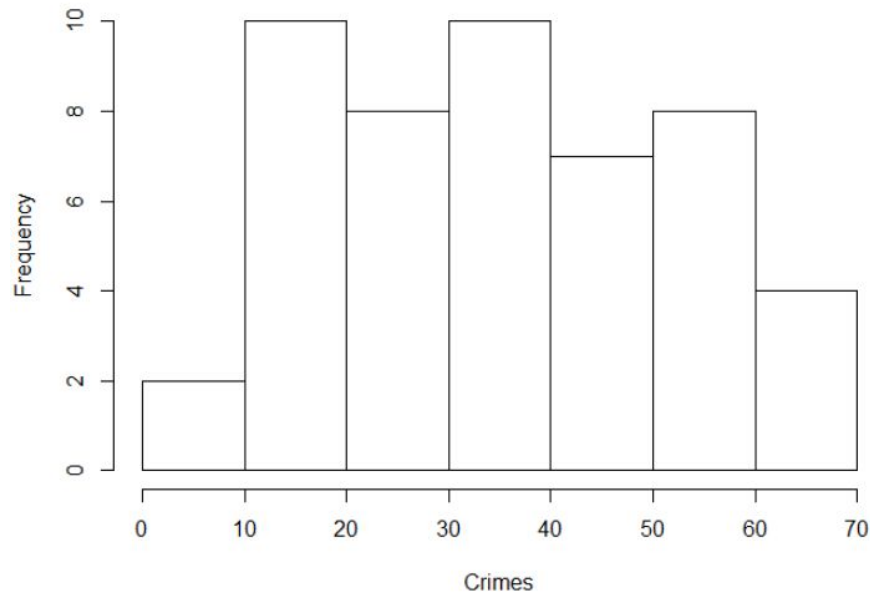
Resultados

Análise exploratória dos dados



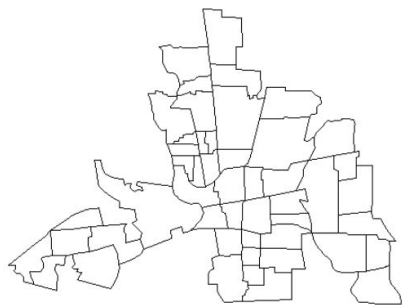
Columbus

Roubos residenciais e furtos de veículos

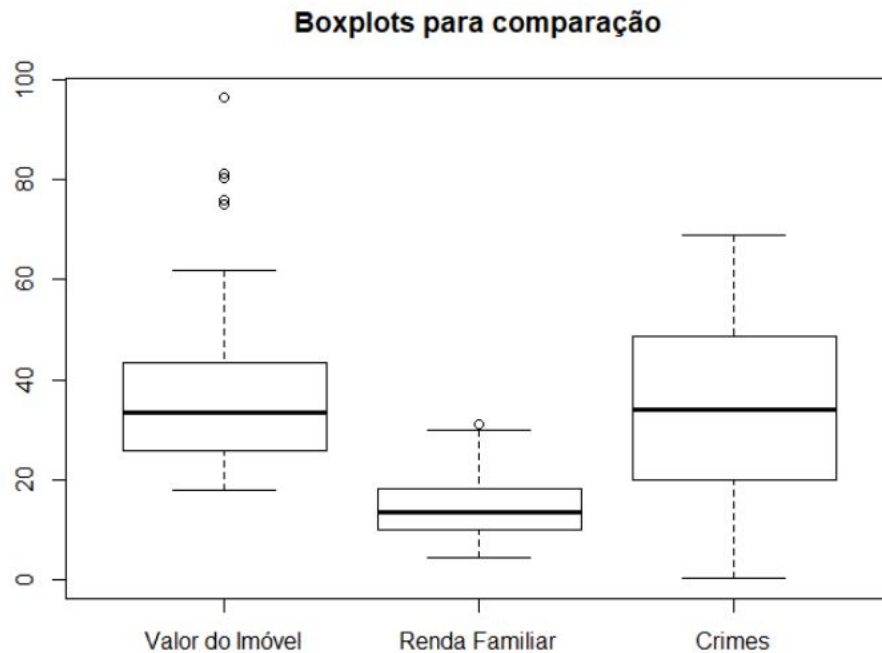


Resultados

Análise exploratória dos dados



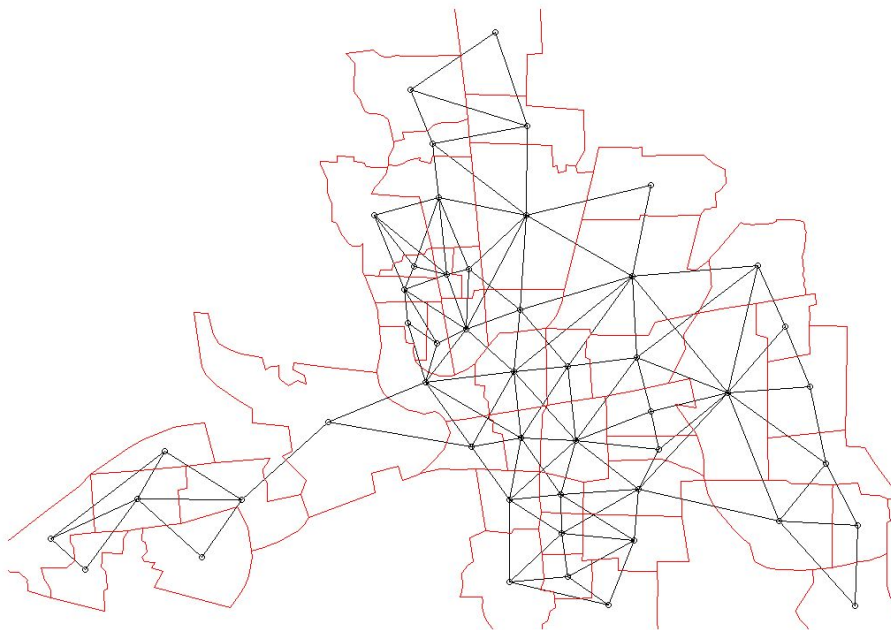
Columbus



Resultados

Identificação dos vizinhos e ligações entre eles:

- 49 regiões formadas;
- 236 ligações;
- Média: 4,816

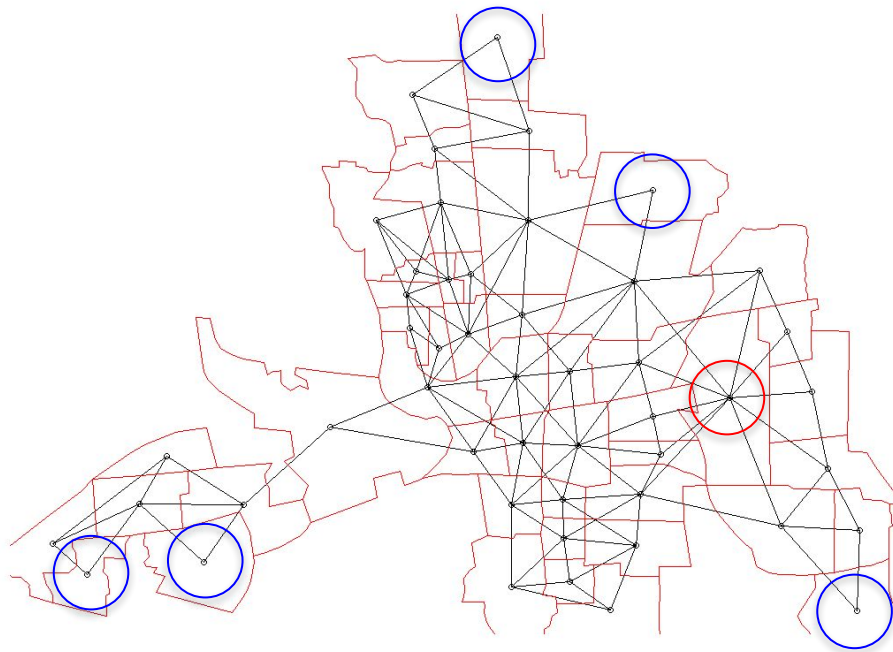


Resultados

Identificação dos vizinhos
e ligações entre eles:

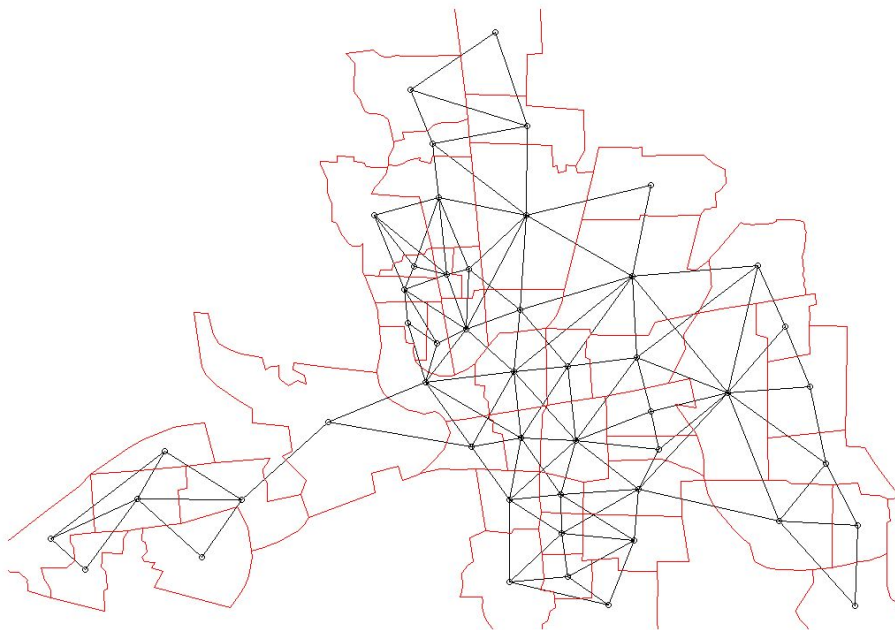
5 regiões com 2 ligações;

1 região com 19 ligações.



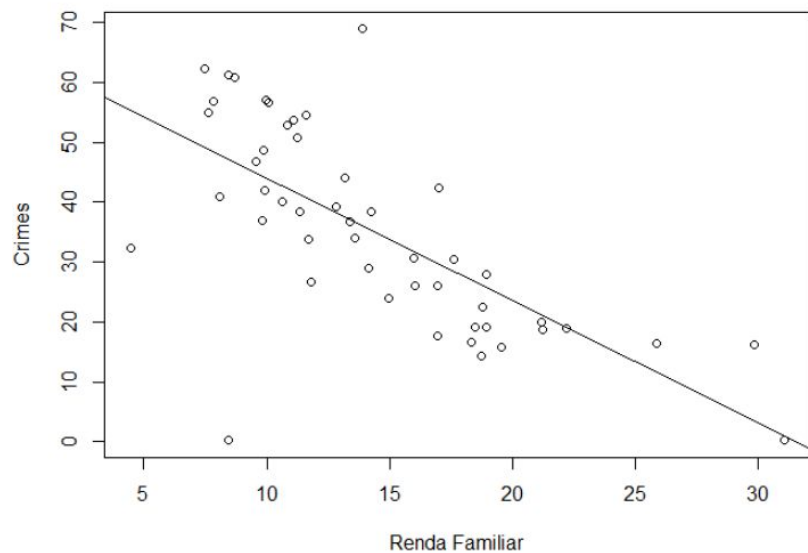
Resultados

Os pesos foram dados a partir do número de ligações com vizinhos, i. e., se há 4 conexões, cada uma receberá 0.25 de peso.

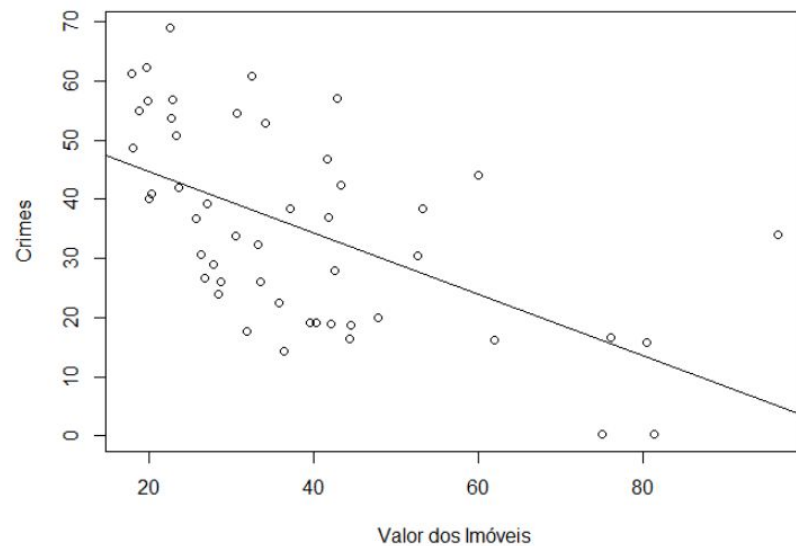


Resultados

Relação entre as variáveis

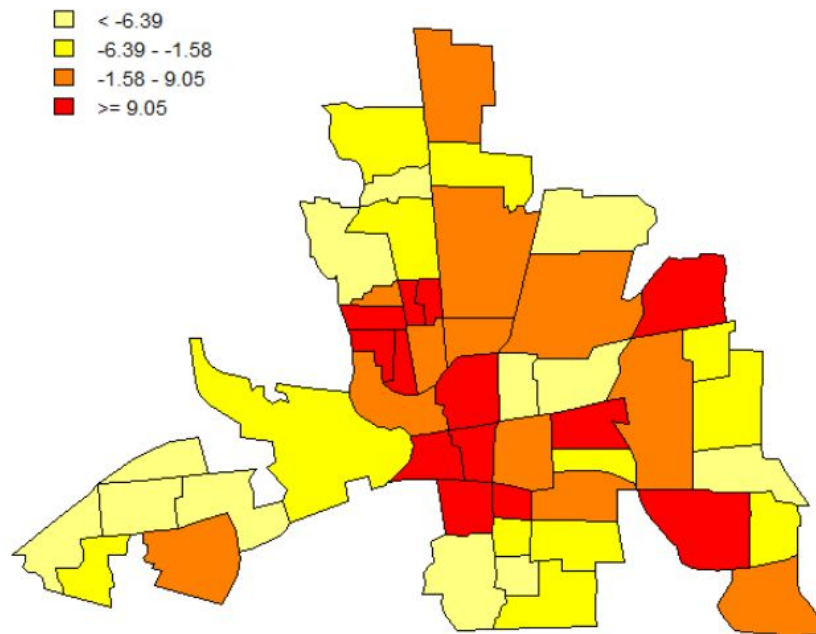
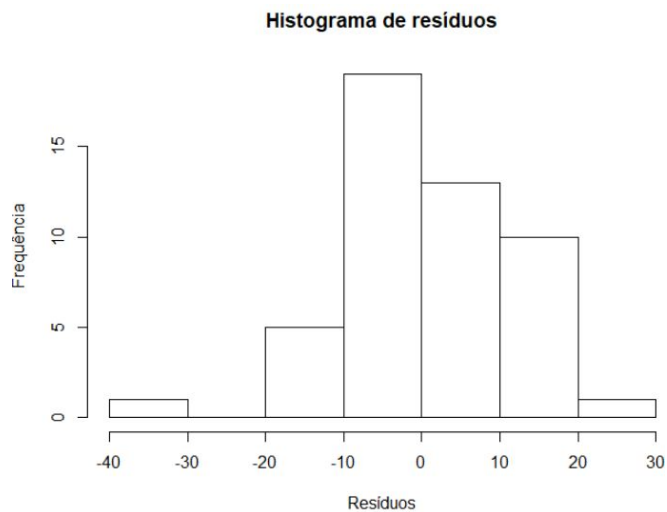


Relação entre as variáveis



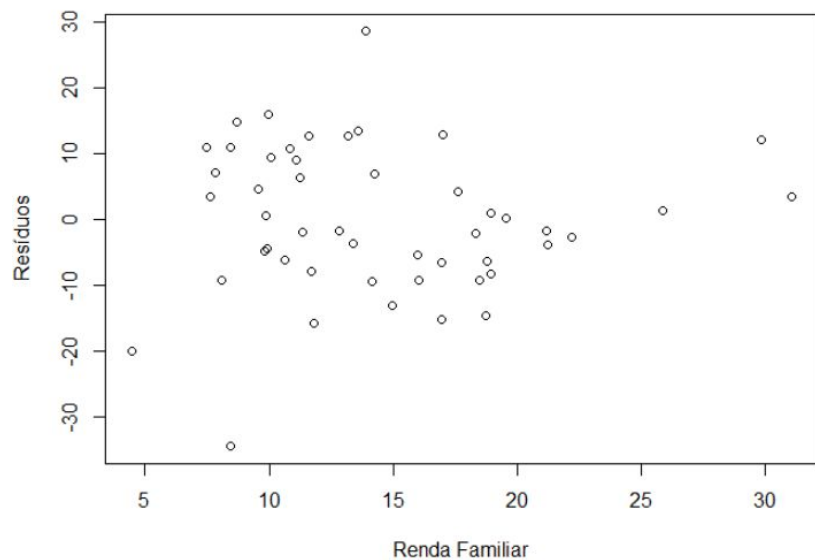
Resultados

Mapa dos resíduos da regressão linear múltipla (CRIME~INC + HOVAL)

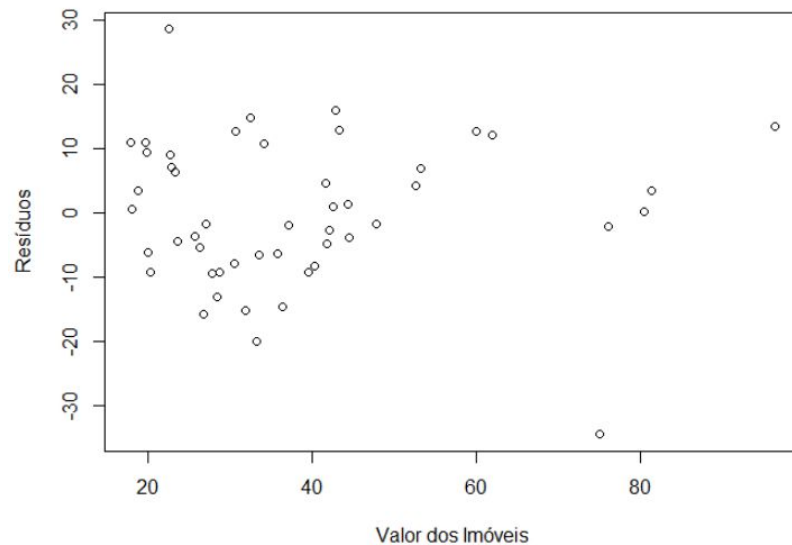


Resultados

Mapa dos resíduos x Renda familiar

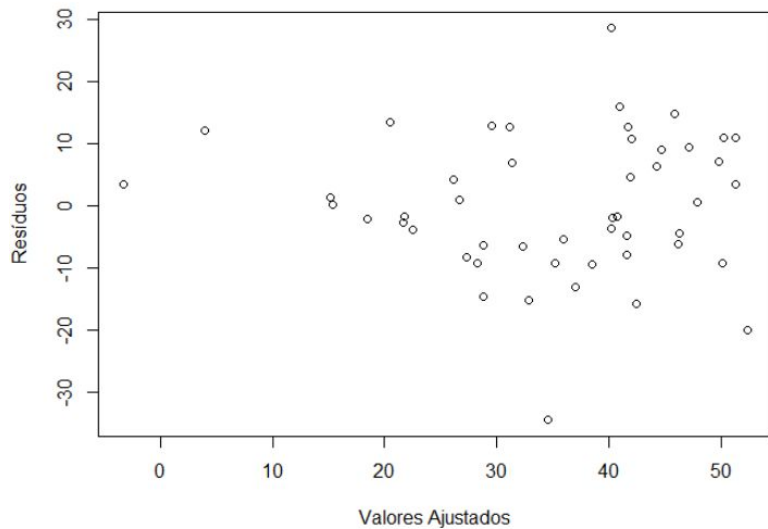


Mapa dos resíduos x Valor dos imóveis

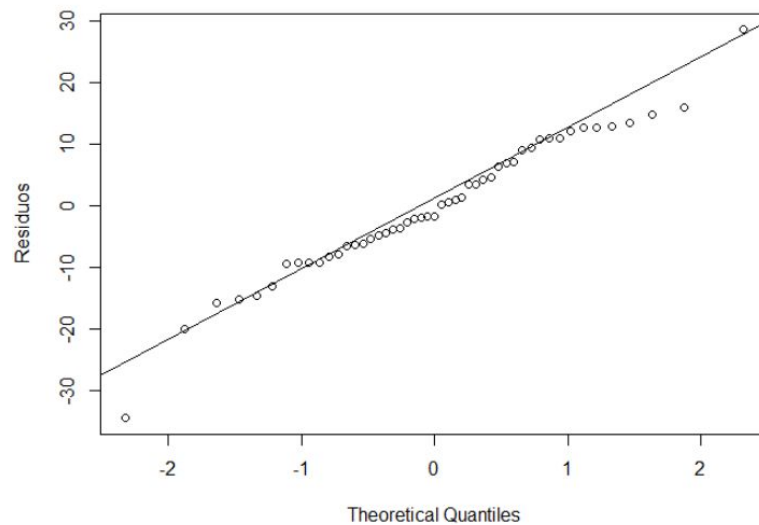


Resultados

Gráfico dos valores ajustados X residuos



Normal Q-Q Plot



Resultados

Testes:

De normalidade:

Jarque Bera Test

```
data: columbus.lm$residuals  
X-squared = 1.8358, df = 2, p-value = 0.3994
```

De heterocedasticidade

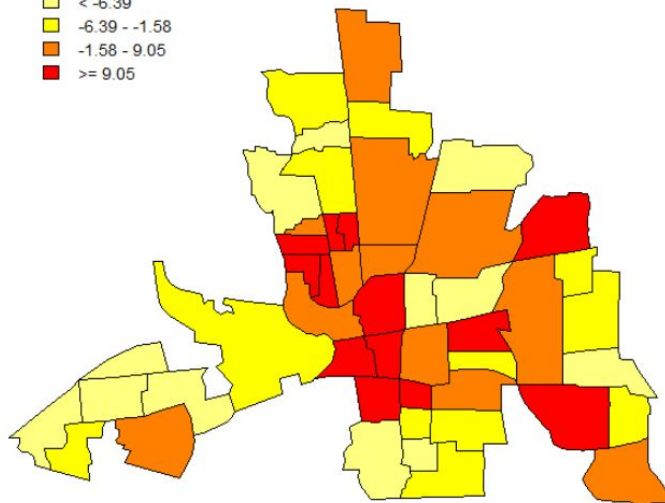
studentized Breusch-Pagan test

```
data: CRIME ~ INC + HOVAL  
BP = 7.2166, df = 2, p-value = 0.0271
```

Resultados

Verificação de autocorrelação espacial:

Mapa dos resíduos da regressão linear múltipla (CRIME-INC + HOVAL)



Global Moran I for regression residuals

```
data:  
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)  
weights: W_polgal
```

```
Moran I statistic standard deviate = 2.8393, p-value = 0.004521  
alternative hypothesis: two.sided  
sample estimates:
```

Observed Moran I	Expectation	Variance
0.222109407	-0.033418335	0.008099305

Resultados

Verificação de dependência espacial:

```
Lagrange multiplier diagnostics for spatial dependence  
data:  
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)  
weights: W_polgal
```

LMerr = 5.2062, df = 1, p-value = 0.02251

```
Lagrange multiplier diagnostics for spatial dependence  
data:  
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)  
weights: W_polgal
```

RLMerr = 0.043906, df = 1, p-value = 0.834

```
Lagrange multiplier diagnostics for spatial dependence  
data:  
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)  
weights: W_polgal
```

LMlag = 8.898, df = 1, p-value = 0.002855

```
Lagrange multiplier diagnostics for spatial dependence  
data:  
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)  
weights: W_polgal
```

RLMlag = 3.7357, df = 1, p-value = 0.05326

Resultados

Verificação de autocorrelação espacial:

```
> summary(columbus.lag)

Call:lagsarlm(formula = CRIME ~ INC + HOVAL, data = columbus, listw = W_polgal)

Residuals:
    Min       1Q   Median       3Q      Max
-37.652017  -5.334611   0.071473   6.107196  23.302618

Type: lag
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  45.603250   7.257404   6.2837 3.306e-10
INC          -1.048728   0.307406  -3.4115 0.000646
HOVAL        -0.266335   0.089096  -2.9893 0.002796

Rho: 0.42333, LR test value: 9.4065, p-value: 0.0021621
Asymptotic standard error: 0.11951
      z-value: 3.5422, p-value: 0.00039686
Wald statistic: 12.547, p-value: 0.00039686

Log likelihood: -182.674 for lag model
ML residual variance (sigma squared): 96.857, (sigma: 9.8416)
Number of observations: 49
Number of parameters estimated: 5
AIC: 375.35, (AIC for lm: 382.75)
LM test for residual autocorrelation
test value: 0.24703, p-value: 0.61917
```

Resultados

Comparação entre os modelos do tipo SAR (Spatial Autoregressive Model) e do tipo CAR (Conditional Autoregressive Model)

```
> summary(columbus.lag)

Call:lagsarlm(formula = CRIME ~ INC + HOVAL, data = columbus, listw = W_polga1)

Residuals:
    Min       1Q   Median       3Q      Max
-37.652017  -5.334611   0.071473   6.107196  23.302618

Type: lag
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 45.603250   7.257404   6.2837 3.306e-10
INC         -1.048728   0.307406  -3.4115 0.000646
HOVAL       -0.266335   0.089096  -2.9893 0.002796

Rho: 0.42333, LR test value: 9.4065, p-value: 0.0021621
Asymptotic standard error: 0.11951
      z-value: 3.5422, p-value: 0.00039686
Wald statistic: 12.547, p-value: 0.00039686

Log likelihood: -182.674 for lag model
ML residual variance (sigma squared): 96.857, (sigma: 9.8416)
Number of observations: 49
Number of parameters estimated: 5
AIC: 375.35, (AIC for lm: 382.75)
LM test for residual autocorrelation
test value: 0.24703, p-value: 0.61917
```

```
> summary(columbus.err)

Call:errorsarlm(formula = CRIME ~ INC + HOVAL, data = columbus, listw = W_polga1)

Residuals:
    Min       1Q   Median       3Q      Max
-34.65998  -6.16943  -0.70623   7.75392  23.43878

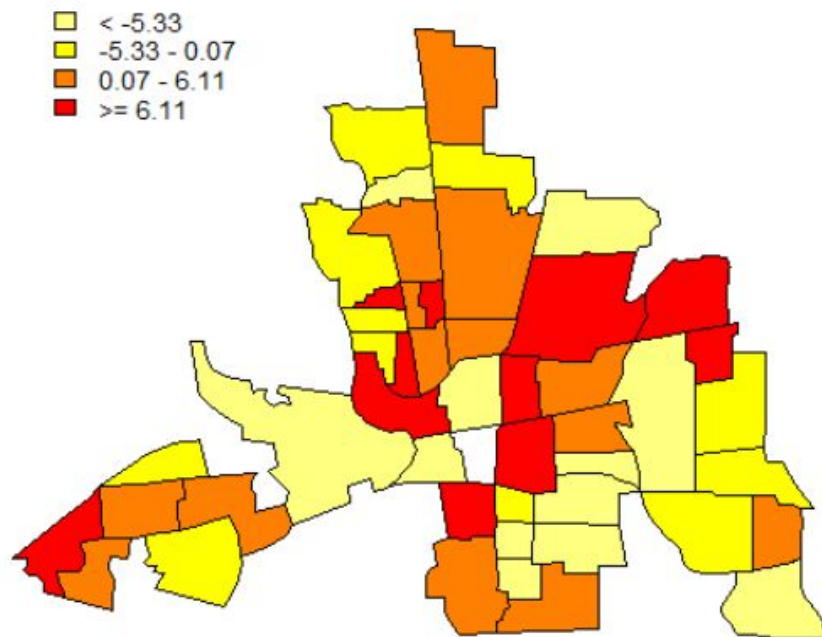
Type: error
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 60.279469   5.365594  11.2344 < 2.2e-16
INC         -0.957305   0.334231  -2.8642 0.0041806
HOVAL       -0.304559   0.092047  -3.3087 0.0009372

Lambda: 0.54675, LR test value: 7.2556, p-value: 0.0070679
Asymptotic standard error: 0.13805
      z-value: 3.9605, p-value: 7.4786e-05
Wald statistic: 15.686, p-value: 7.4786e-05

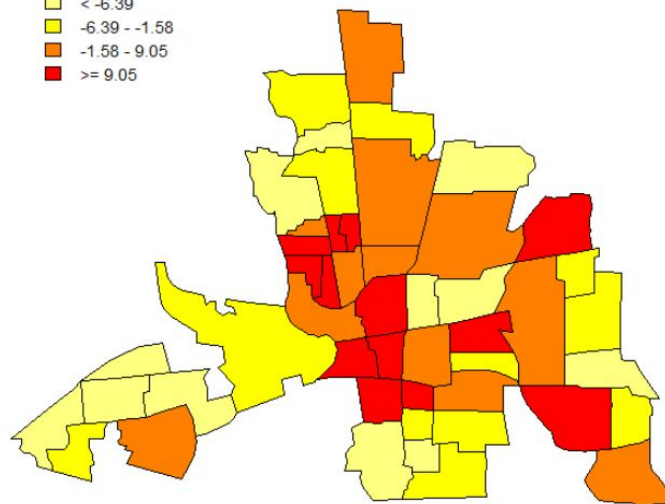
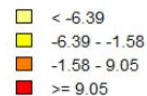
Log likelihood: -183.7494 for error model
ML residual variance (sigma squared): 97.674, (sigma: 9.883)
Number of observations: 49
Number of parameters estimated: 5
AIC: 377.5, (AIC for lm: 382.75)
```

Resultados

Mapa dos resíduos do modelo SAR : CRIME-INC+HOVAL

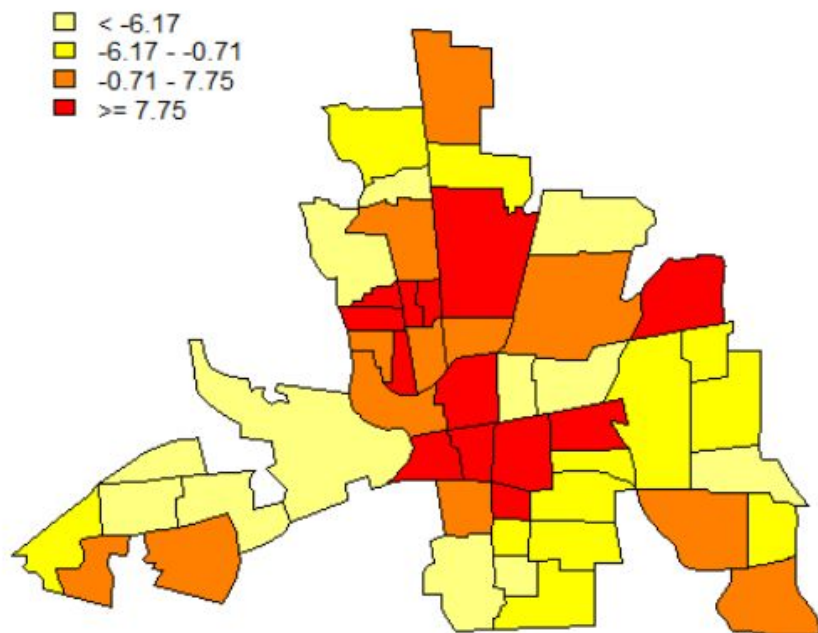


Mapa dos resíduos da regressão linear múltipla (CRIME-INC + HOVAL)

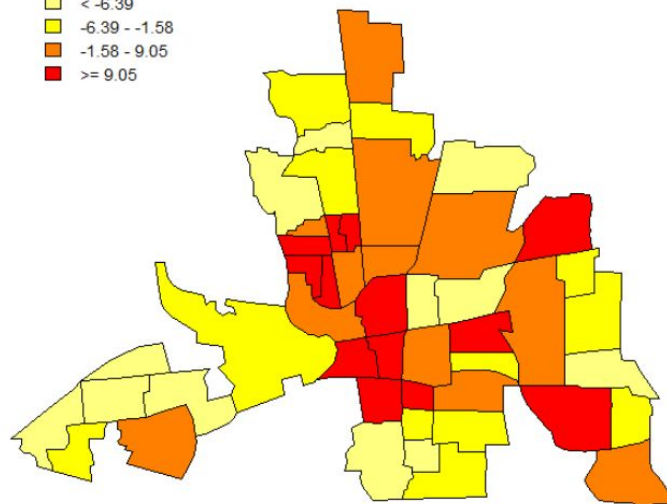
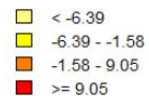


Resultados

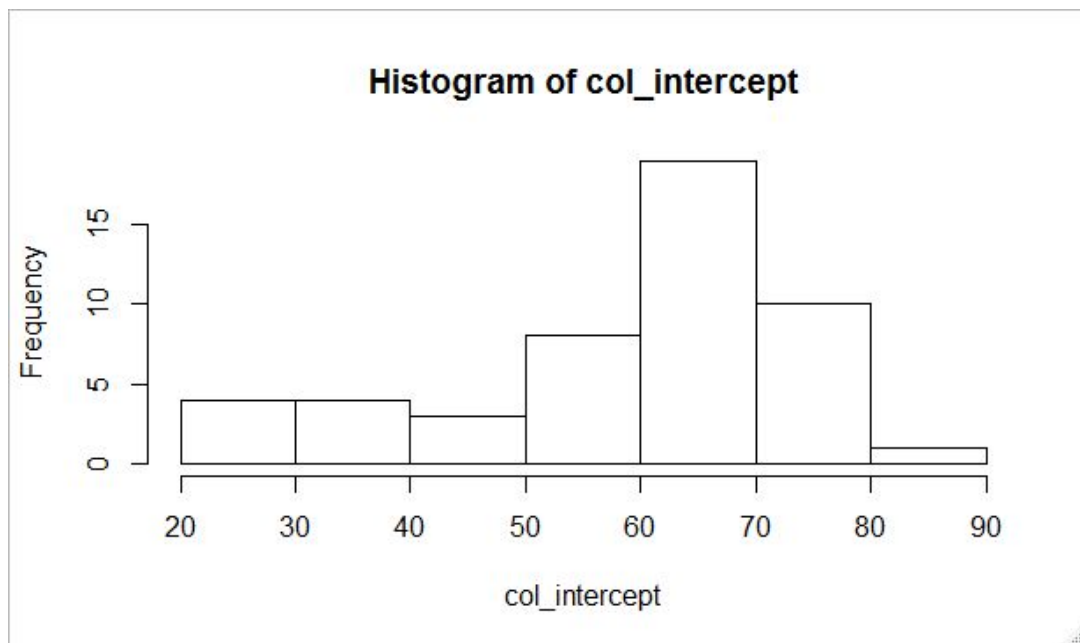
Mapa dos resíduos do modelo CAR : CRIME~INC+HOVAL



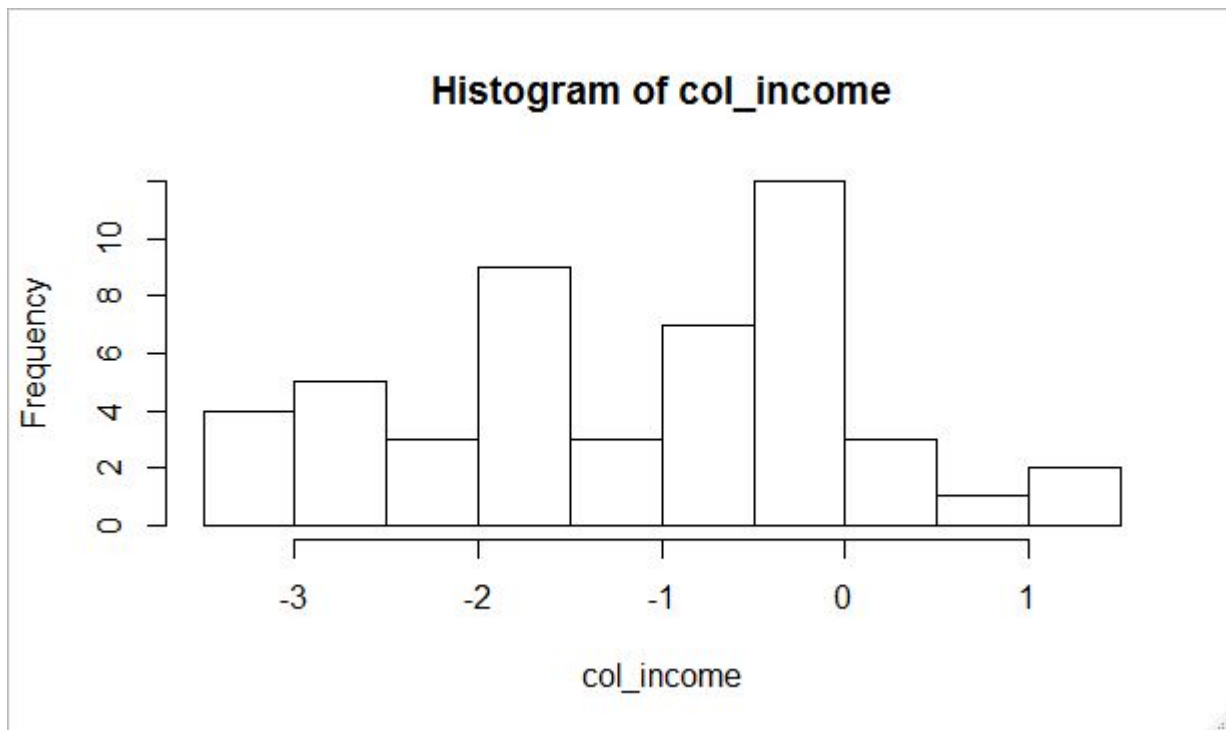
Mapa dos resíduos da regressão linear múltipla (CRIME~INC + HOVAL)



Resultados



Resultados



Conclusões

Análise da correlação e distribuição espacial dos dados;

Admite-se a dependência espacial dos dados, analisados em função das distâncias (matrizes de vizinhança) e variabilidade espacial;

Avalia-se a correlação espacial dos dados a partir da análise de tendências;

Valores extremos qualificam a significância dos dados analisados em função do problema.

Os testes estatísticos nos permitem avaliar os valores com real significância.