ANÁLISE DE REGRESSÃO ESPACIAL

Curso: Análise Espacial de Dados Geográficos

Docentes: Antônio Miguel Vieira Monteiro e Eduardo G. Camargo

Discentes: Beatriz Davida da Silva; Danylo Magalhães; Gustavo Piva Lopes Salgado

REGRESSÃO LINEAR SIMPLES

$$Yi=\beta 0+\beta 1.Xi+\varepsilon i$$

Y = variável resposta (dependente)

Xi = variável preditora (independente)

β0 = intercepto populacional (coeficiente de regressão)

β1 = inclinação populacional (coeficiente de regressão)

ε = erro aleatório (Y real – Y estimado)

REGRESSÃO LINEAR MÚLTIPLA

$$Yi=\beta 0+\beta 1.Xi1+\beta 2.Xi2+\varepsilon i$$

INC (Income)

variável independente (Xi1)

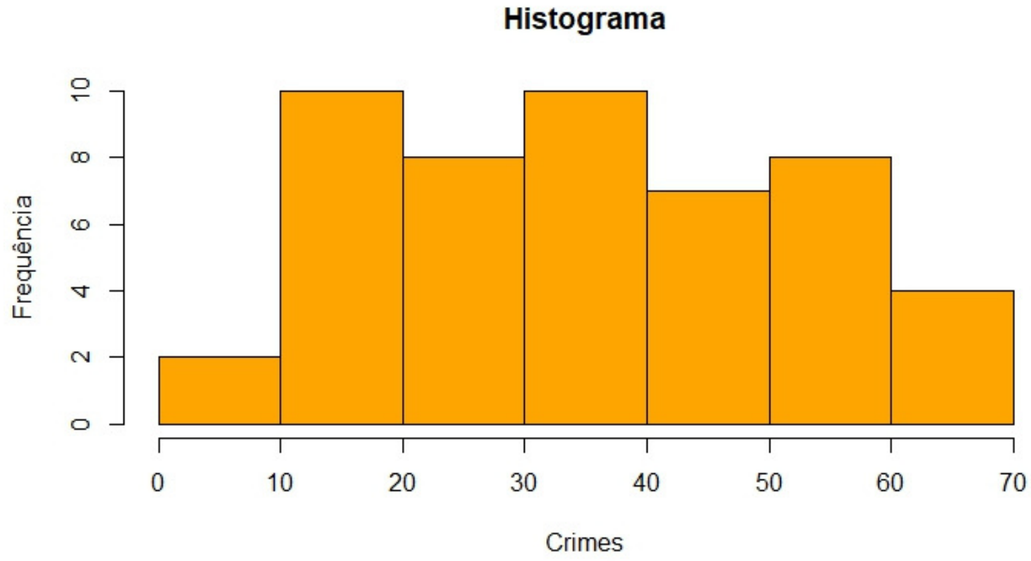
HOVAL (House value)

variável independente (Xi2)

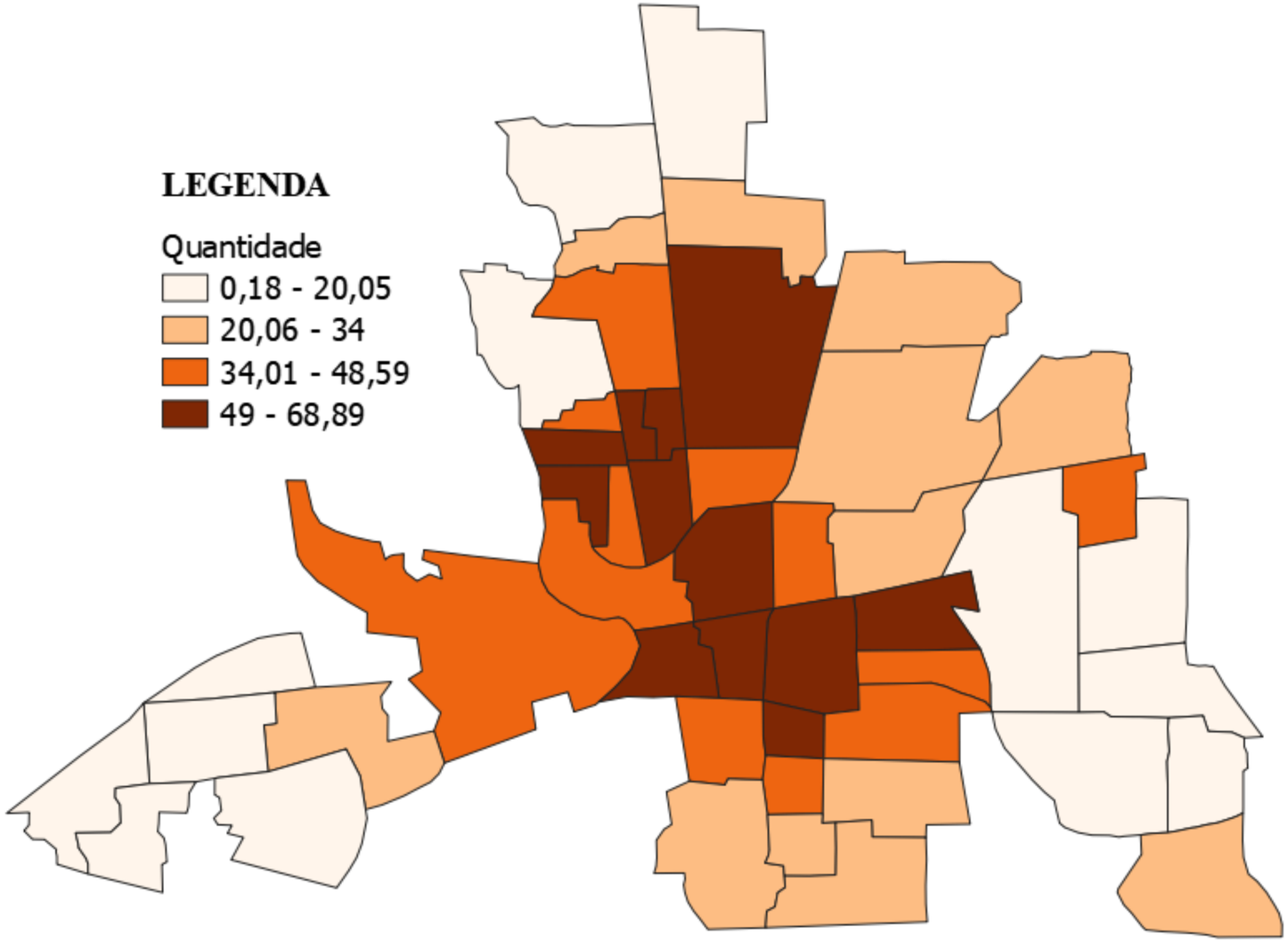
CRIME

variável resposta (Y)

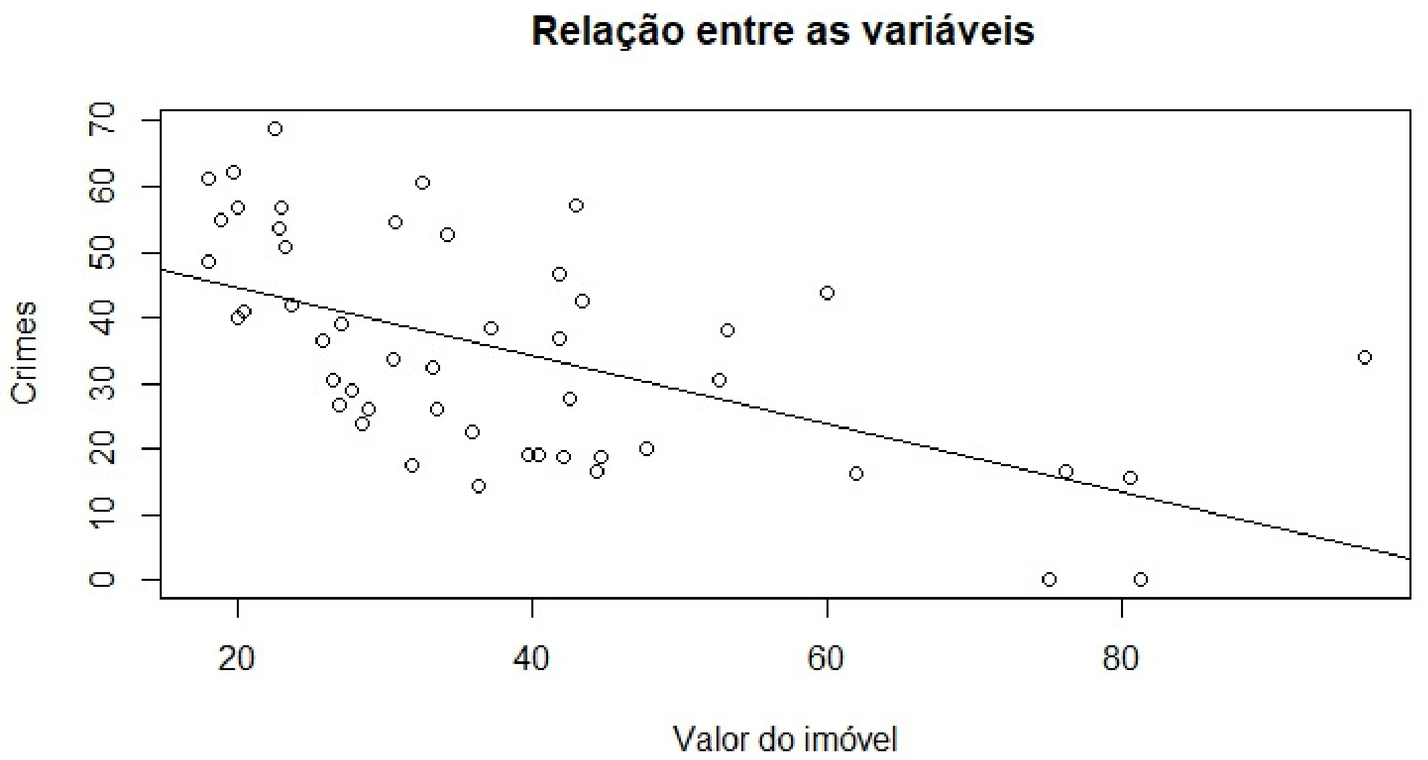
ANÁLISE EXPLORATÓRIA DOS DADOS

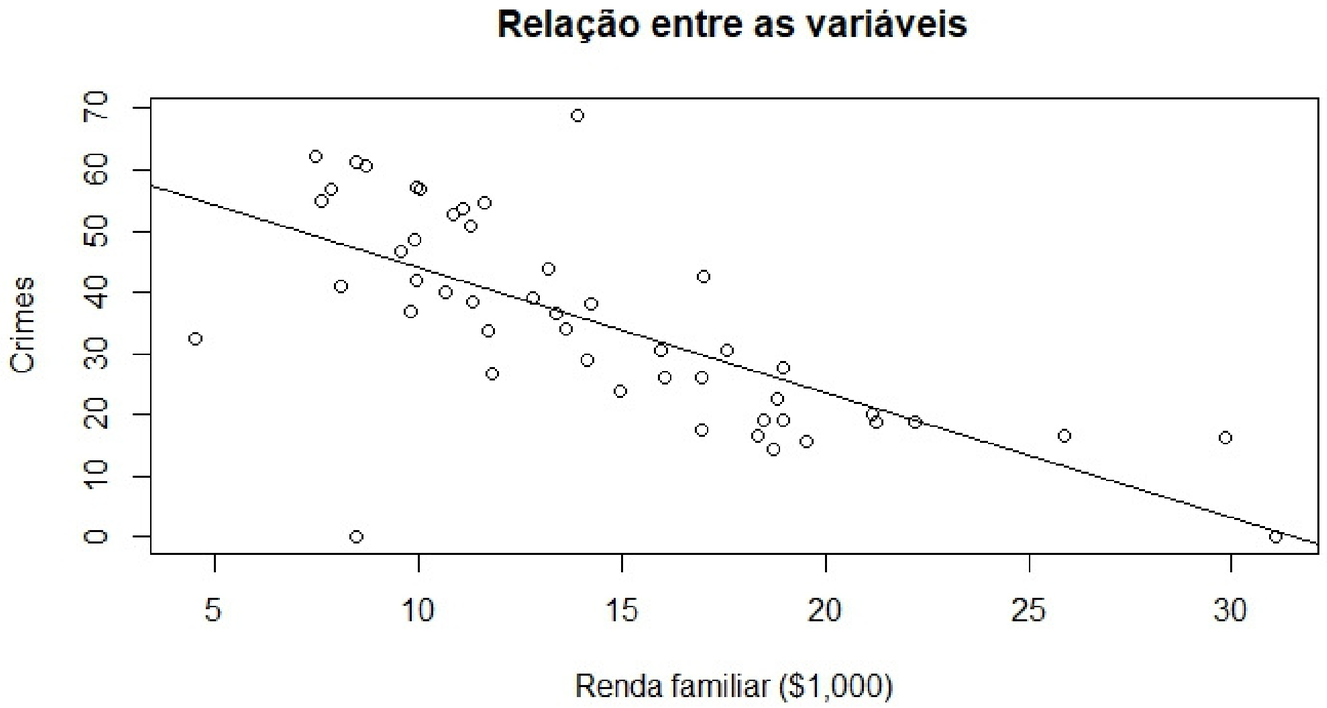


Roubos residenciais e furtos de veículos



ANÁLISE EXPLORATÓRIA DOS DADOS





EQUAÇÃO DE REGRESSÃO LINEAR MÚLTIPLA

```
Call:
lm(formula = CRIME ~ INC + HOVAL, data = columbus)

Residuals:
    Min      1Q  Median      3Q     Max
-34.418  -6.388  -1.580   9.052  28.649

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  68.6190     4.7355  14.490  < 2e-16 ***
INC          -1.5973     0.3341  -4.780 1.83e-05 ***
HOVAL        -0.2739     0.1032  -2.654   0.0109 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.43 on 46 degrees of freedom
Multiple R-squared:  0.5524, Adjusted R-squared:  0.5329
F-statistic: 28.39 on 2 and 46 DF,  p-value: 9.341e-09
```

Crimes em Columbus

O modelo de regressão linear múltipla para o conjunto de dados é representado pela equação:

$$\text{CRIME} = 68{,}6190 - 1{,}5973 \times \text{INC} - 0{,}2739 \times \text{HOVAL}$$

Todos os três coeficientes (Intercepto, INC e HOVAL) têm valores-p muito baixos, indicando que são estatisticamente significativos.

PREMISSAS PARA REGRESSÃO LINEAR ADEQUADA

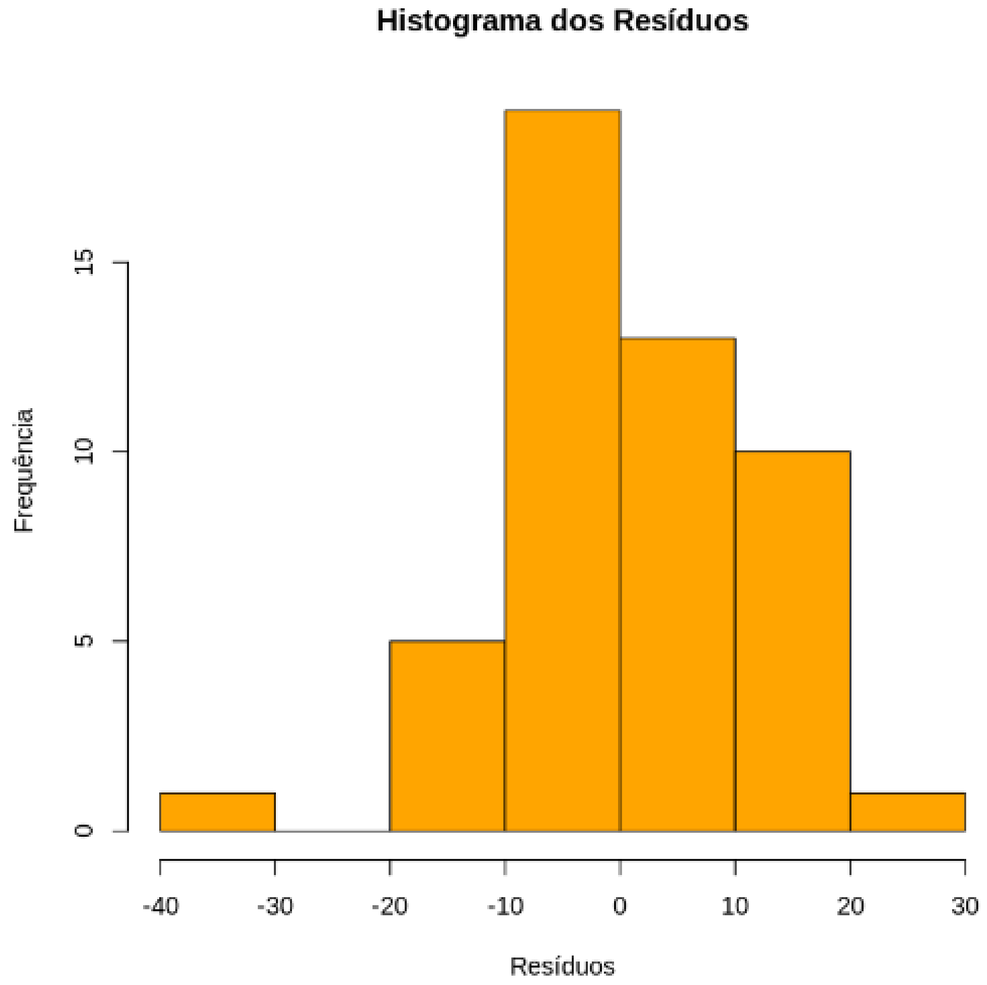
1. **Distribuição Normal** da variável resposta Y;
2. **Linearidade** de todos os valores médios de Y;
3. **Homocedasticidade**, com a variância de Y sendo a mesma para qualquer valor de X;
4. **Independência** dos valores de Yi e Yj.

NORMALIDADE DA VARIÁVEL RESPOSTA Y

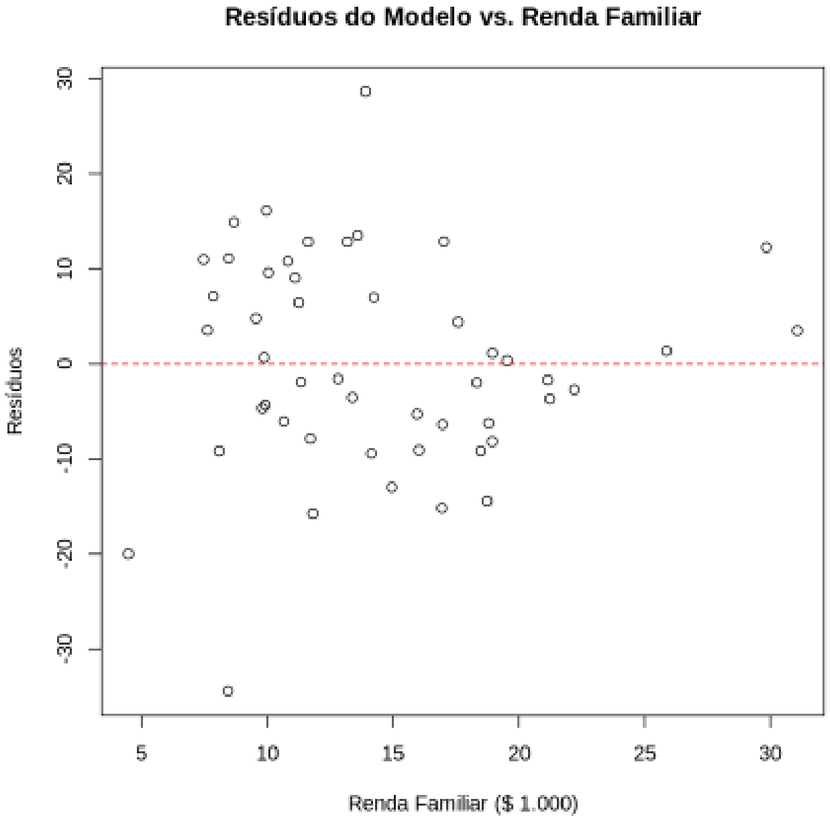
Teste de normalidade Shapiro-Wilk

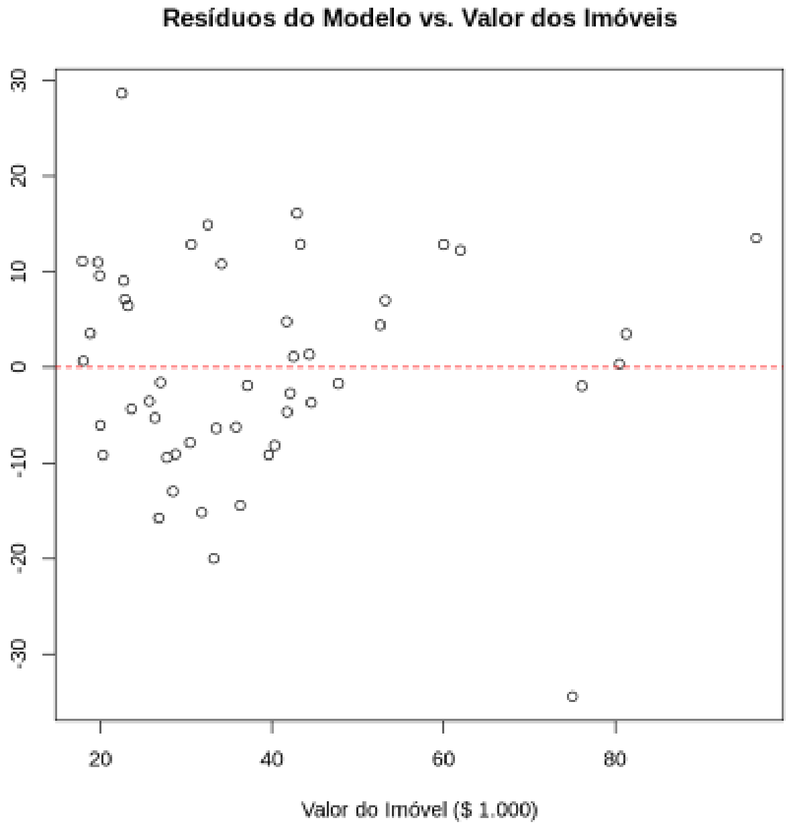
```
        Shapiro-Wilk normality test

data:  residuos
W = 0.97708, p-value = 0.4497
```

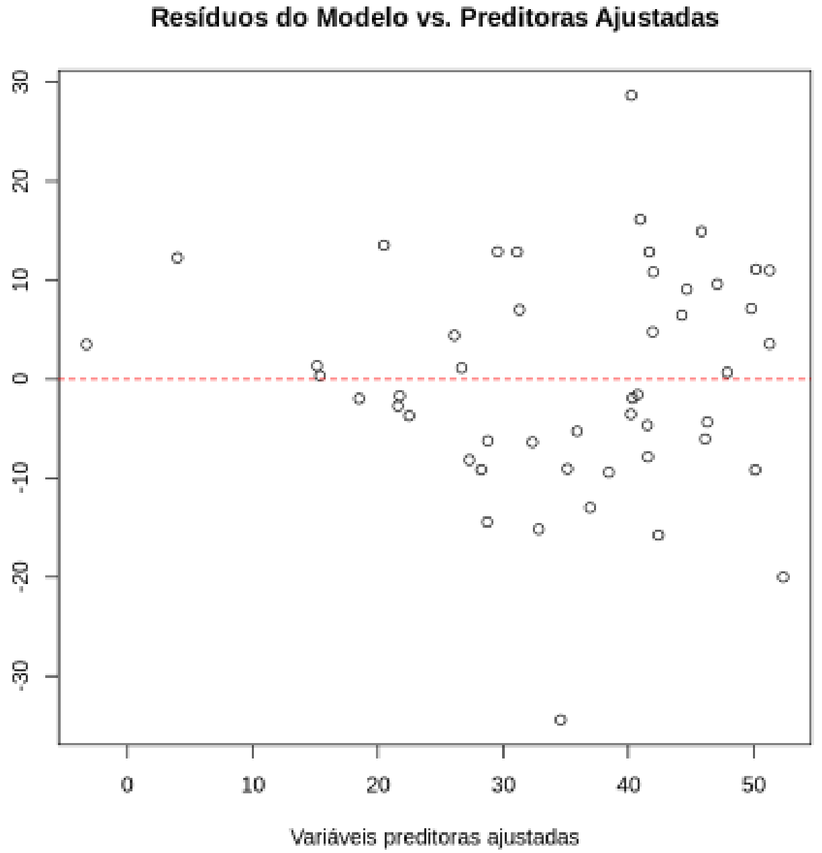
Valor-p maior que 5% (0,05)



LINEARIDADE DO MODELO







LINEARIDADE DO MODELO

```
Call:
lm(formula = CRIME ~ INC + HOVAL, data = columbus)

Residuals:
    Min      1Q  Median      3Q     Max
-34.418  -6.388  -1.580   9.052  28.649

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  68.6190     4.7355  14.490  < 2e-16 ***
INC          -1.5973     0.3341  -4.780 1.83e-05 ***
HOVAL        -0.2739     0.1032  -2.654   0.0109 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.43 on 46 degrees of freedom
Multiple R-squared:  0.5524,    Adjusted R-squared:  0.5329
F-statistic: 28.39 on 2 and 46 DF,  p-value: 9.341e-09
```

R^2 Múltiplo: O modelo explica 55,24% da variância na variável resposta Y.

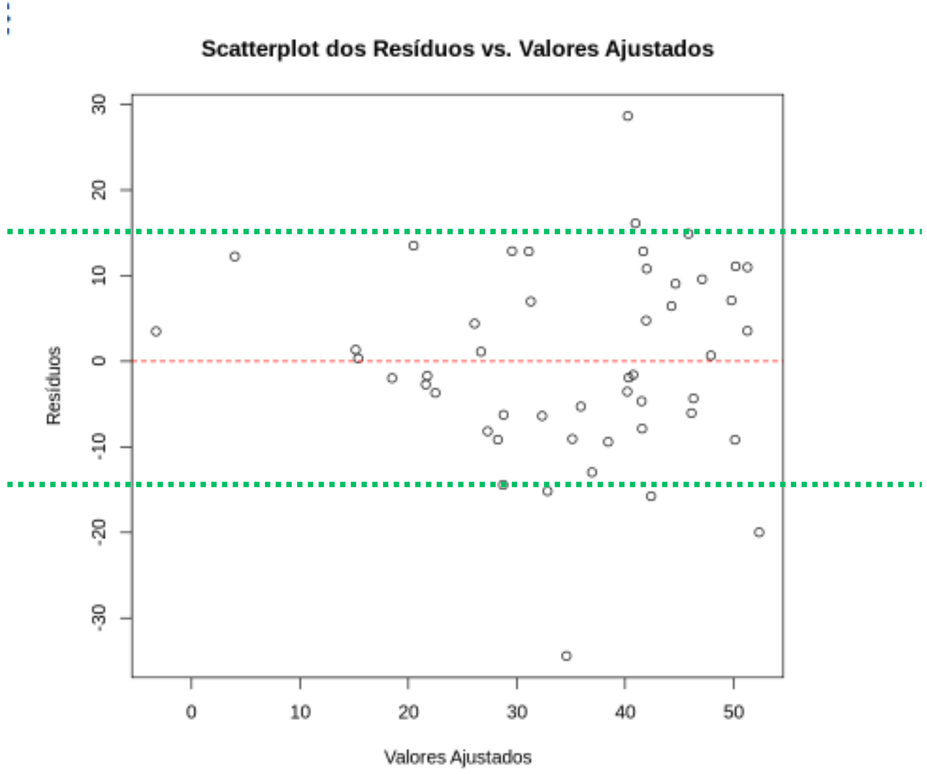
R^2 Ajustado: O modelo, ajustado aos números de preditores, explica 53,29% da variância na variável resposta Y.

HOMOCEDASTICIDADE / VARIÂNCIA DE Y CONSTANTE

Teste de Homocedasticidade de Breusch-Pagan

```
        studentized Breusch-Pagan test

data:  columbus.lm
BP = 7.2166, df = 2, p-value = 0.0271
```

Valor-p menor que 5% (0,05)



INDEPENDÊNCIA ESPACIAL - ÍNDICE DE MORAN

```
        Global Moran I for regression residuals

data:
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
weights: col.listw

Moran I statistic standard deviate = 2.681, p-value = 0.00734
alternative hypothesis: two.sided
sample estimates:
Observed Moran I      Expectation         Variance
     0.212374153     -0.033268284      0.008394853
```

Autocorrelação espacial constatada!!!

teste é significativo a 5% de significância

-1 : significa autocorrelação espacial negativa ou inversa;

0 : significa aleatoriedade;

+1 : significa **autocorrelação espacial positiva** ou direta.

MULTIPLICADOR DE LAGRANGE

```
        Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
weights: col.listw

LMerr = 4.6111, df = 1, p-value = 0.03177
```

Primeiro é analisado o Valor-P dos testes LMerr e LMlag

Há evidências de autocorrelação espacial

Os dois testes são significativos

```
        Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
weights: col.listw

LMlag = 7.8557, df = 1, p-value = 0.005066
```

Há evidências de autocorrelação espacial

MULTIPLICADOR DE LAGRANGE

```
        Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
weights: col.listw

RLMerr = 0.033514, df = 1, p-value = 0.8547
```

Como os dois testes anteriores são significativos, é preciso analisar os testes robustos

Não significativo, valor-P maior que 0,05

```
        Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
weights: col.listw

RLMlag = 3.2781, df = 1, p-value = 0.07021
```

Esse é o modelo que deve ser utilizado

Spatial Lag Model (SAR)

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

Residuals:
       Min         1Q     Median         3Q        Max
-37.4497093  -5.4565567   0.0016387   6.7159553  24.7107978

Type: lag
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 46.851431   7.314754  6.4051 1.503e-10
INC         -1.073533   0.310872 -3.4533 0.0005538
HOVAL       -0.269997   0.090128 -2.9957 0.0027381

Rho: 0.40389, LR test value: 8.4179, p-value: 0.0037154
Asymptotic standard error: 0.12071
    z-value: 3.3459, p-value: 0.00082027
Wald statistic: 11.195, p-value: 0.00082027

Log likelihood: -183.1683 for lag model
ML residual variance (sigma squared): 99.164, (sigma: 9.9581)
Number of observations: 49
Number of parameters estimated: 5
AIC: 376.34, (AIC for lm: 382.75)
LM test for residual autocorrelation
test value: 0.19184, p-value: 0.66139
```

Parâmetro autoregressivo espacial (Rho) é altamente significativo, indicado pelo valor-P (0.0008) do teste assintótico (grandes amostras)

Logaritmo de Verossimilhança (LR) também é altamente significativo (valor-p 0,0037).

Spatial Error Model (CAR)

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

Residuals:
      Min        1Q    Median        3Q       Max
-34.45950  -6.21730  -0.69775   7.65256  24.23631

Type: error
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 61.053618   5.314875 11.4873 < 2.2e-16
INC         -0.995473   0.337025 -2.9537 0.0031398
HOVAL       -0.307979   0.092584 -3.3265 0.0008794

Lambda: 0.52089, LR test value: 6.4441, p-value: 0.011132
Asymptotic standard error: 0.14129
    z-value: 3.6868, p-value: 0.00022713
Wald statistic: 13.592, p-value: 0.00022713

Log likelihood: -184.1552 for error model
ML residual variance (sigma squared): 99.98, (sigma: 9.999)
Number of observations: 49
Number of parameters estimated: 5
AIC: NA (not available for weighted model), (AIC for lm: 382.75)
```

Ao comparar os dois modelos pelo valor do log da verossimilhança, nota-se que a do Modelo SAR é melhor.

Geographically Weighted Regression (GWR)

```
Call:
gwr(formula = CRIME ~ INC + HOVAL, data = columbus, coords = cbind(columbus$Y,
    columbus$X), adapt = gwr_kernel, hatmatrix = TRUE, se.fit = TRUE)
Kernel function: gwr.Gauss
Adaptive quantile: 0.1349225 (about 6 of 49 data points)
Summary of GWR coefficient estimates at data points:
                  Min.   1st Qu.    Median   3rd Qu.      Max.  Global
X.Intercept. 61.045867 66.107912 68.865429 72.046543 75.823831 68.6190
INC          -2.746460 -1.979271 -1.817207 -0.995802  0.745660 -1.5973
HOVAL        -0.734161 -0.380004 -0.232414 -0.077966  0.180549 -0.2739
Number of data points: 49
Effective number of parameters (residual: 2traceS - traceS'S): 13.93087
Effective degrees of freedom (residual: 2traceS - traceS'S): 35.06913
Sigma (residual: 2traceS - traceS'S): 9.861483
Effective number of parameters (model: traceS): 10.52469
Effective degrees of freedom (model: traceS): 38.47531
Sigma (model: traceS): 9.414856
Sigma (ML): 8.342702
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 377.9159
AIC (GWR p. 96, eq. 4.22): 357.4766
Residual sum of squares: 3410.433
Quasi-global R2: 0.7462139
```

Isso indica que as relações entre as variáveis são locais e podem diferir espacialmente.

O desvio padrão dos resíduos indica a variabilidade dos erros em diferentes locais.

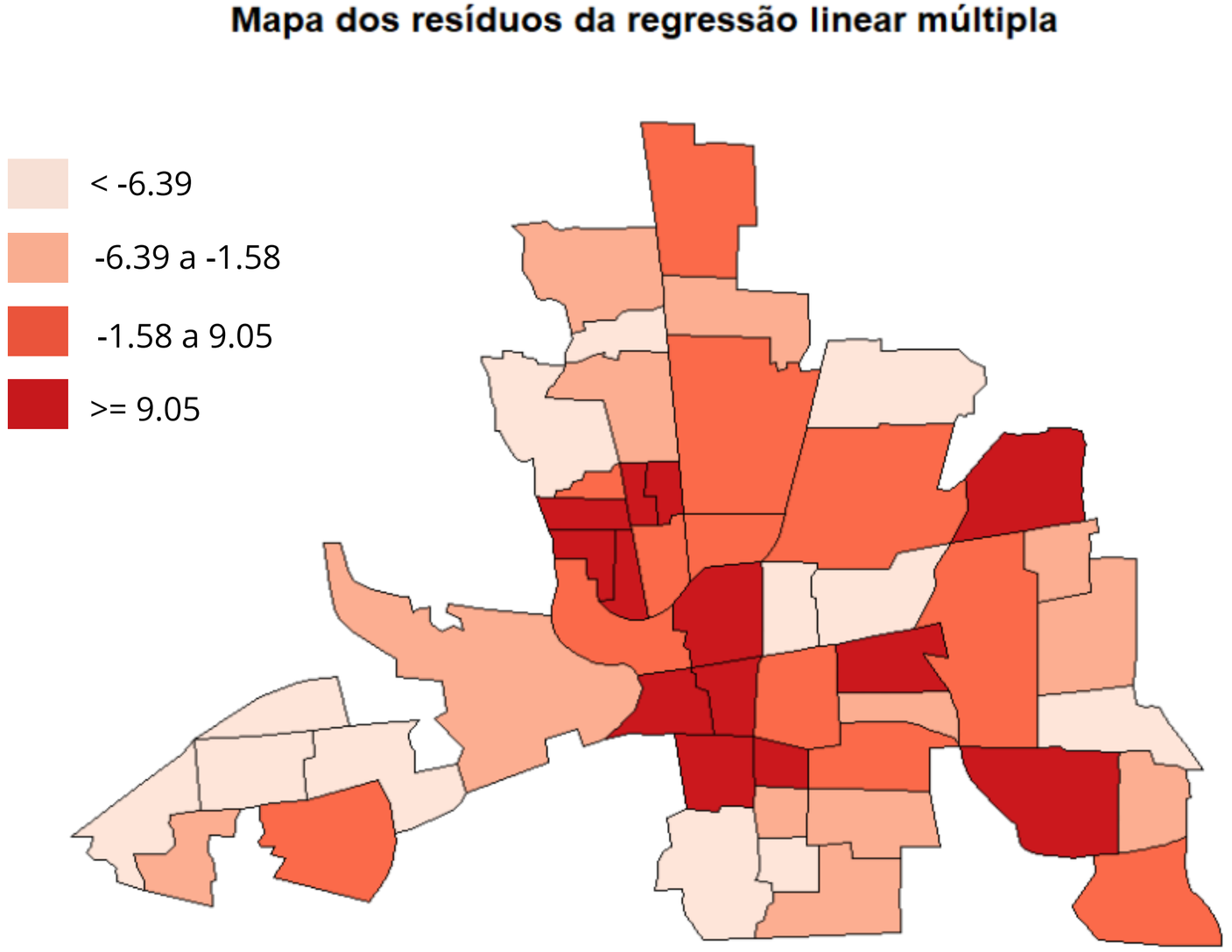
AIC do SAR é menor, logo por esse critério o SAR é melhor.

Valor maior que o Modelo Linear comum, o que indica um ajuste melhor do Modelo GWR

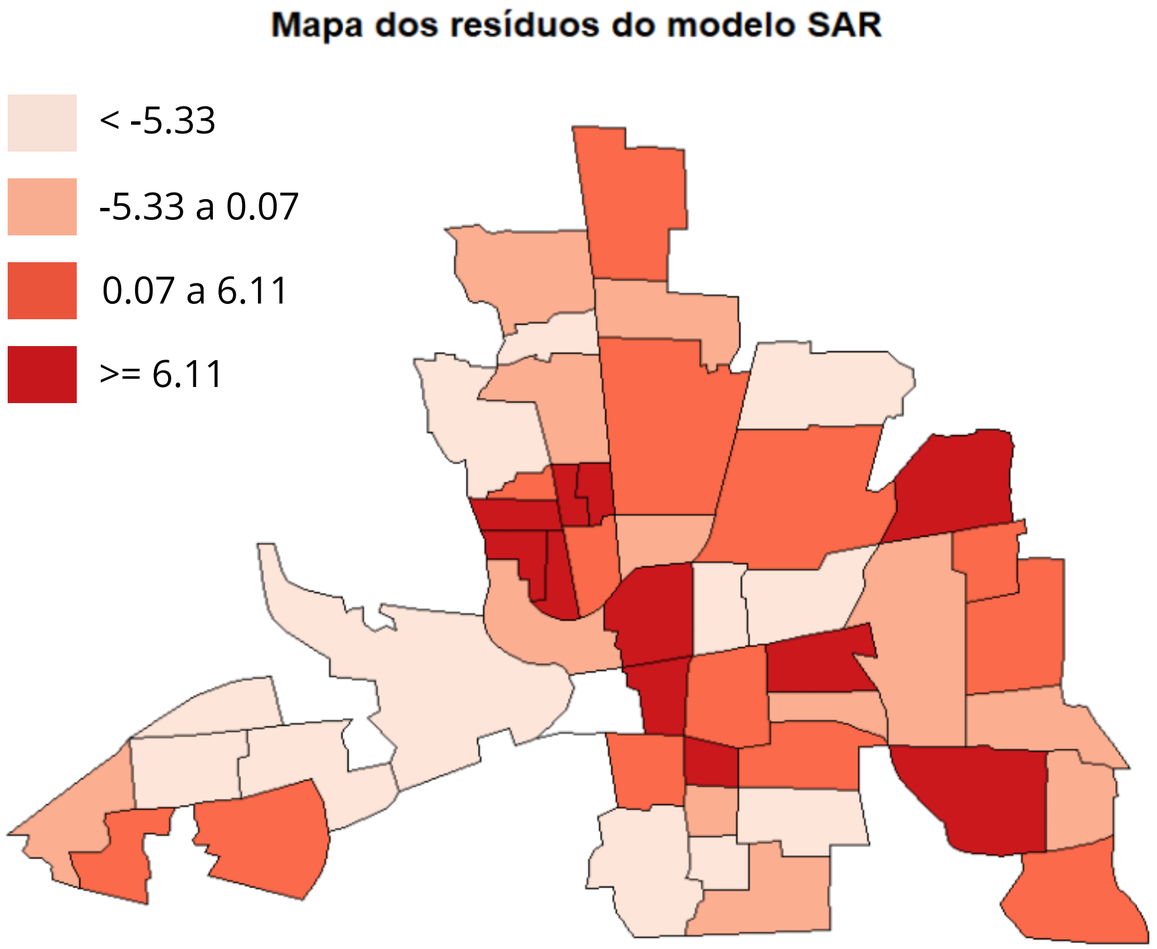
COMPARAÇÕES DAS REGRESSÕES

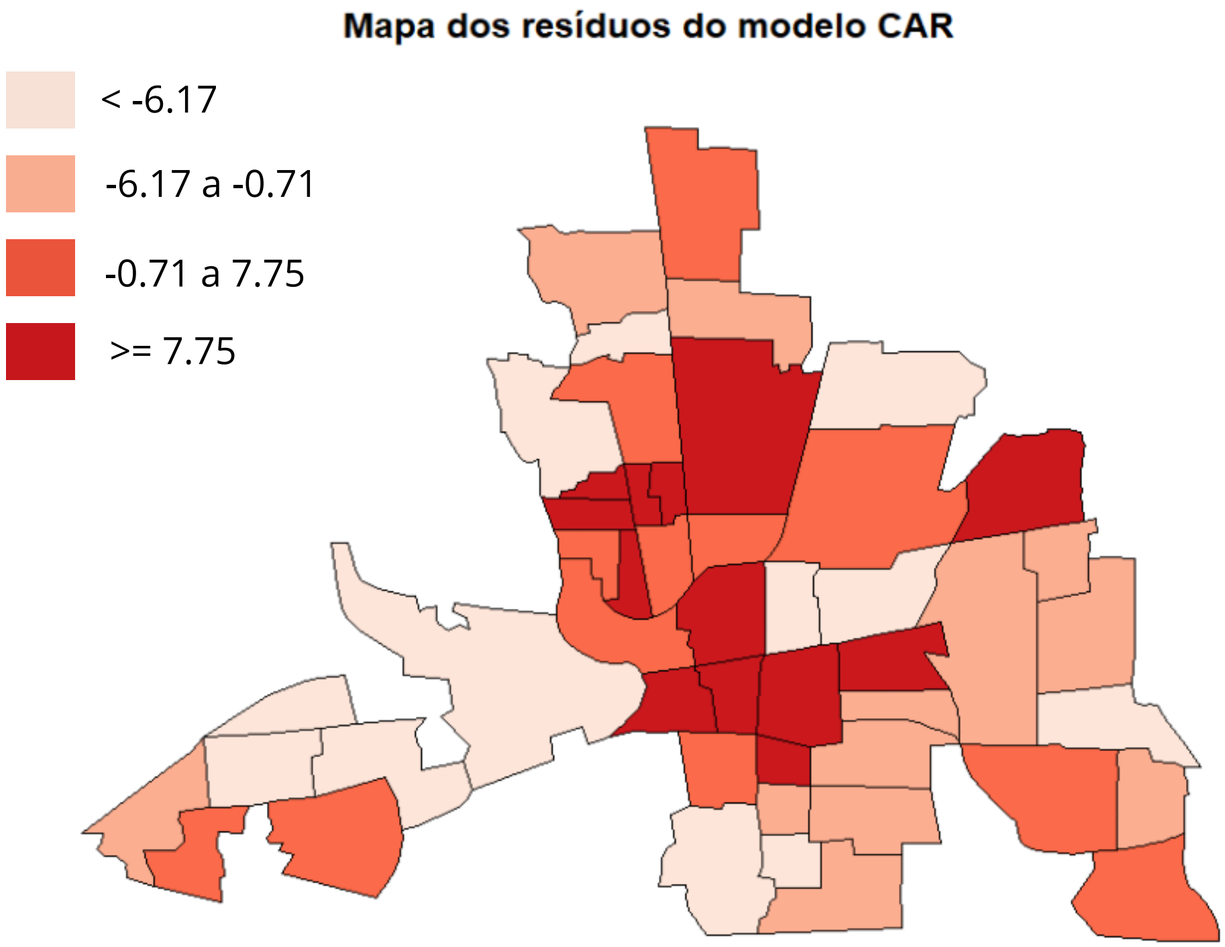
	Regressão Linear Múltipla	SAR	GWR
R^2 ajustado	0,53	0,61	0,75
AIC	382,75	376,34	357,48
Indice Moran dos resíduos	0,21	0.029376269	0.02043468

MAPA DOS RESÍDUOS DA REGRESSÃO LINEAR MÚLTIPLA

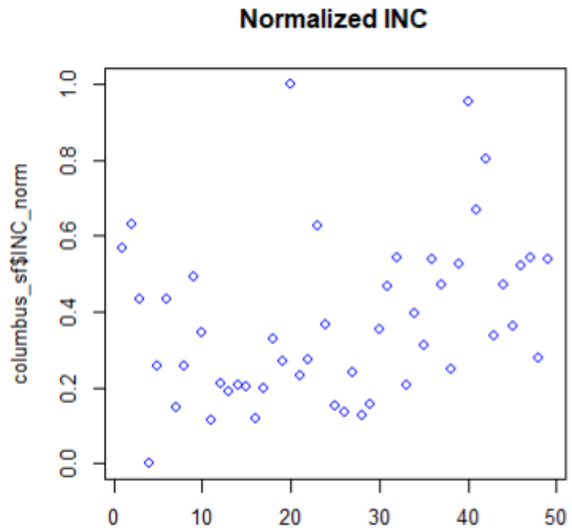


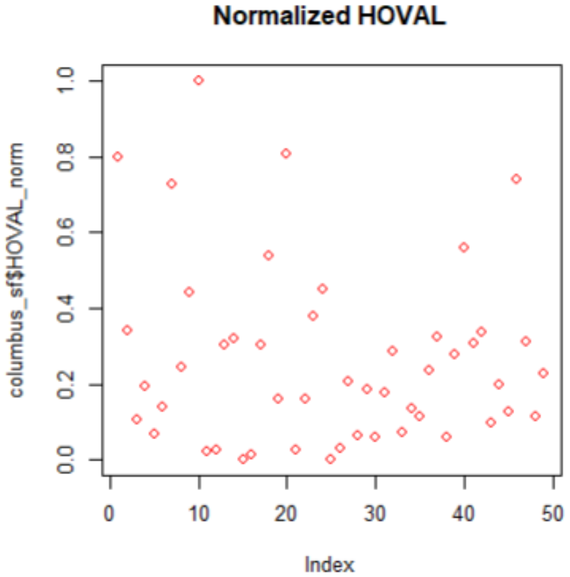
MAPA DOS RESÍDUOS - SAR e CAR

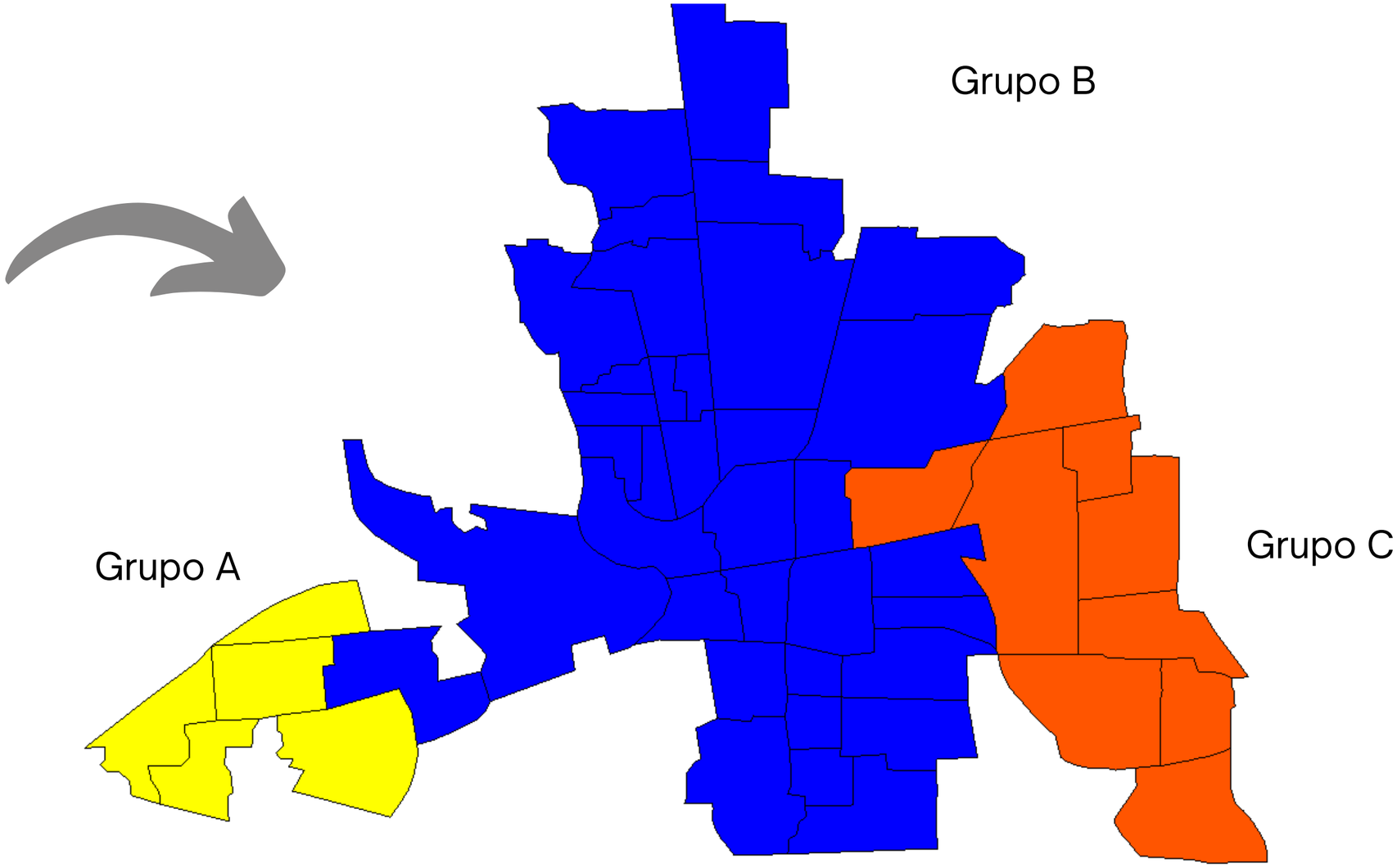




REGIONALIZAÇÃO - SKATER







Spatial Lag Model (SAR) - Região A

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

Residuals:
     Min       1Q   Median       3Q      Max
-1.92612 -0.14462  0.10750  0.20337  1.75987

Type: lag
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 37.0121595  6.1647493  6.0038 1.927e-09
INC         -0.3310859  0.1671771 -1.9804   0.04765
HOVAL       -0.0076855  0.0333318 -0.2306   0.81764

Rho: -0.81932, LR test value: 1.4439, p-value: 0.2295
Asymptotic standard error: 0.326
    z-value: -2.5132, p-value: 0.011963
Wald statistic: 6.3164, p-value: 0.011963

Log likelihood: -8.55876 for lag model
ML residual variance (sigma squared): 1.3762, (sigma: 1.1731)
Number of observations: 5
Number of parameters estimated: 5
AIC: 27.118, (AIC for lm: 26.561)
LM test for residual autocorrelation
test value: 2.9893, p-value: 0.083817
```

Spatial Error Model (CAR) - Região A

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

Residuals:
     Min       1Q   Median       3Q      Max
-0.93085 -0.25930 -0.23664  0.27413  1.15267

Type: error
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 26.387358   2.190720 12.0451 < 2.2e-16
INC         -0.489187   0.095779 -5.1074 3.266e-07
HOVAL       -0.011113   0.021359 -0.5203    0.6028

Lambda: -1.1865, LR test value: 4.3023, p-value: 0.038061
Asymptotic standard error: 0.13256
    z-value: -8.9502, p-value: < 2.22e-16
Wald statistic: 80.105, p-value: < 2.22e-16

Log likelihood: -7.129576 for error model
ML residual variance (sigma squared): 0.47871, (sigma: 0.69189)
Number of observations: 5
Number of parameters estimated: 5
AIC: NA (not available for weighted model), (AIC for lm: 26.561)
```

Spatial Lag Model (SAR) - Região B

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

Residuals:
      Min        1Q    Median        3Q       Max
-28.27797  -7.76056  -0.86774   7.74240  21.31884

Type: lag
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 43.63913    9.67366  4.5111 6.448e-06
INC         -0.74876    0.49740 -1.5054 0.1322325
HOVAL       -0.45324    0.12624 -3.5902 0.0003304

Rho: 0.5053, LR test value: 6.1139, p-value: 0.013412
Asymptotic standard error: 0.15506
    z-value: 3.2586, p-value: 0.0011195
Wald statistic: 10.619, p-value: 0.0011195

Log likelihood: -132.6092 for lag model
ML residual variance (sigma squared): 106.94, (sigma: 10.341)
Number of observations: 35
Number of parameters estimated: 5
AIC: 275.22, (AIC for lm: 279.33)
LM test for residual autocorrelation
test value: 1.6768, p-value: 0.19536
```

Spatial Error Model (CAR) - Região B

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

Residuals:
      Min        1Q    Median        3Q       Max
-26.01790  -7.91034  -0.35789   7.07228  19.38909

Type: error
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 61.43561    7.80449  7.8718 3.553e-15
INC         -0.42918    0.53405 -0.8036    0.4216
HOVAL       -0.50569    0.12301 -4.1109 3.942e-05

Lambda: 0.59596, LR test value: 6.1942, p-value: 0.012817
Asymptotic standard error: 0.1522
    z-value: 3.9156, p-value: 9.0186e-05
Wald statistic: 15.332, p-value: 9.0186e-05

Log likelihood: -132.5691 for error model
ML residual variance (sigma squared): 103.34, (sigma: 10.166)
Number of observations: 35
Number of parameters estimated: 5
AIC: NA (not available for weighted model), (AIC for lm: 279.33)
```

Spatial Lag Model (SAR) - Região C

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

Residuals:
      Min        1Q    Median        3Q       Max
-6.514457 -0.651896  0.027097  1.236884  5.871618

Type: lag
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 56.454010   5.482457 10.2972 < 2.2e-16
INC         -1.432235   0.183343 -7.8118 5.551e-15
HOVAL        0.073047   0.075785  0.9639    0.3351

Rho: -0.47698, LR test value: 1.3159, p-value: 0.25132
Asymptotic standard error: 0.29589
    z-value: -1.612, p-value: 0.10696
Wald statistic: 2.5985, p-value: 0.10696

Log likelihood: -24.68286 for lag model
ML residual variance (sigma squared): 13.245, (sigma: 3.6393)
Number of observations: 9
Number of parameters estimated: 5
AIC: 59.366, (AIC for lm: 58.682)
LM test for residual autocorrelation
test value: 0.020573, p-value: 0.88595
```

Spatial Error Model (CAR) - Região C

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

Residuals:
     Min       1Q   Median       3Q      Max
-6.59143 -1.00290 -0.30489  1.81367  7.64619

Type: error
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) 50.7951691  4.8042741 10.5729 < 2.2e-16
INC         -1.4347934  0.2014143 -7.1236 1.051e-12
HOVAL        0.0078055  0.0695879  0.1122    0.9107

Lambda: -0.06118, LR test value: 0.016012, p-value: 0.89931
Asymptotic standard error: 0.42861
    z-value: -0.14274, p-value: 0.8865
Wald statistic: 0.020374, p-value: 0.8865

Log likelihood: -25.33283 for error model
ML residual variance (sigma squared): 16.29, (sigma: 4.0361)
Number of observations: 9
Number of parameters estimated: 5
AIC: NA (not available for weighted model), (AIC for lm: 58.682)
```

Regionalização Skater

Modelo mais adequado para cada região:

Região A --> CAR

Região B --> CAR

Região C --> Modelo Linear

GWR

GWR4 Semiparametric GWR/GWGL modelling tool

File (F) Help (H)

Step 1: Data > | Step 2: Model > | Step 3: Kernel > | Step 4: Output > | Step 5: Execute

Locational variables
ID key (optional)
005 POLYID
X coordinate (Lon)
013 X
Y coordinate (Lat)
014 Y
Projected Spherical

Model settings
Model type
Gaussian
Poisson (count)
Logistic (binary)

Options
Standardisation of independent variables
Geographical variability test
L -> G variable-selection
G -> L variable-selection

Variable (Field) list
001 AREA
002 PERIMETER
003 COLUMBUS_
004 COLUMBUS_I
006 NEIG
010 OPEN
011 PLUMB
012 DISCBD
015 NSA
016 NSB
017 EW
018 CP
019 THOUS
020 NEIGNO

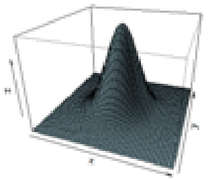
Regression variables
Dependent variable
009 CRIME
Offset variable
(Optional for a Poisson model)
Independent variables
Local (L):
000 Intercept
007 HOVAL
008 INC
Global (G):

GWR4 Semiparametric GWR/GWGL modelling tool

File (F) Help (H)

Step 1: Data > | Step 2: Model > | Step 3: Kernel > | Step 4: Output > | Step 5: Execute

Kernel type
Fixed Gaussian (distance)
Fixed bi-square (distance)
Adaptive bi-square (NN)
Adaptive Gaussian (NN)



Bandwidth selection method
Golden section search
Use user-defined range
Max.
Min.
Single bandwidth
Interval Search
Max.
Min.
Interval

X and Y ranges
calculate

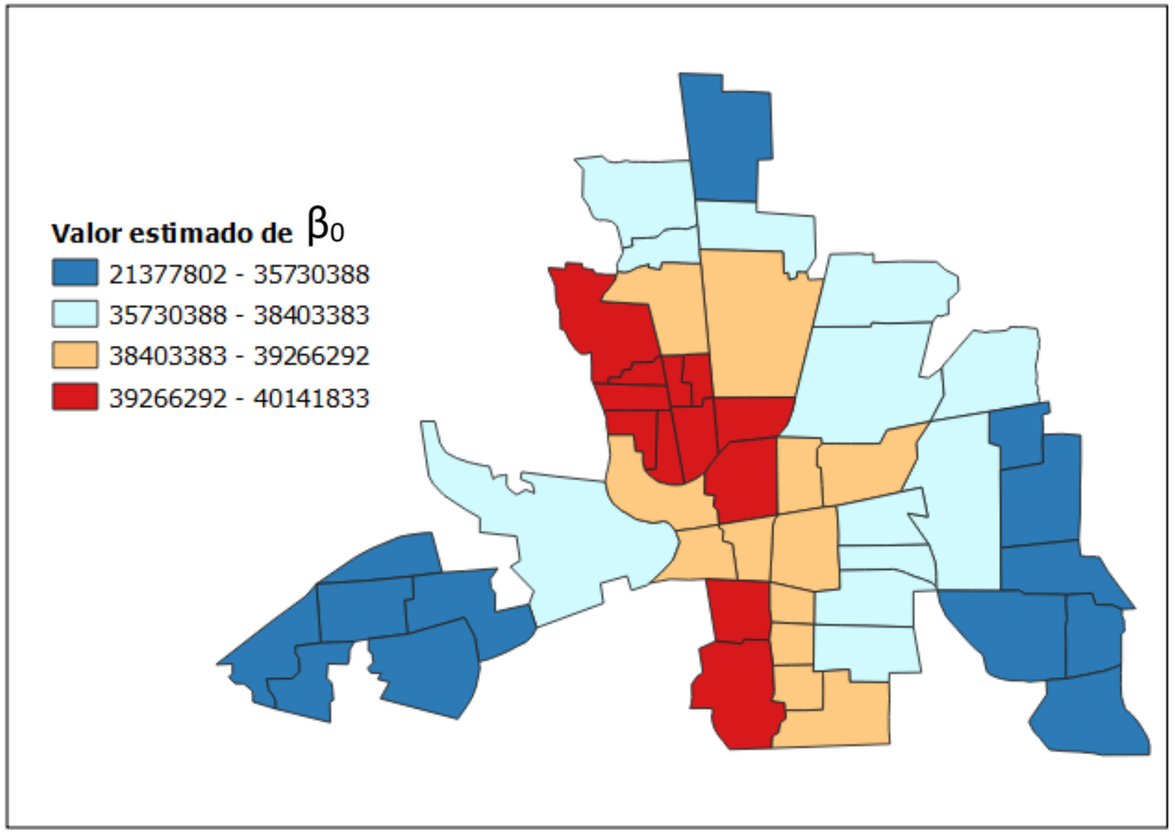
Selection Criteria
AICc
AIC
BIC/MDL
CV (Gaussian only)

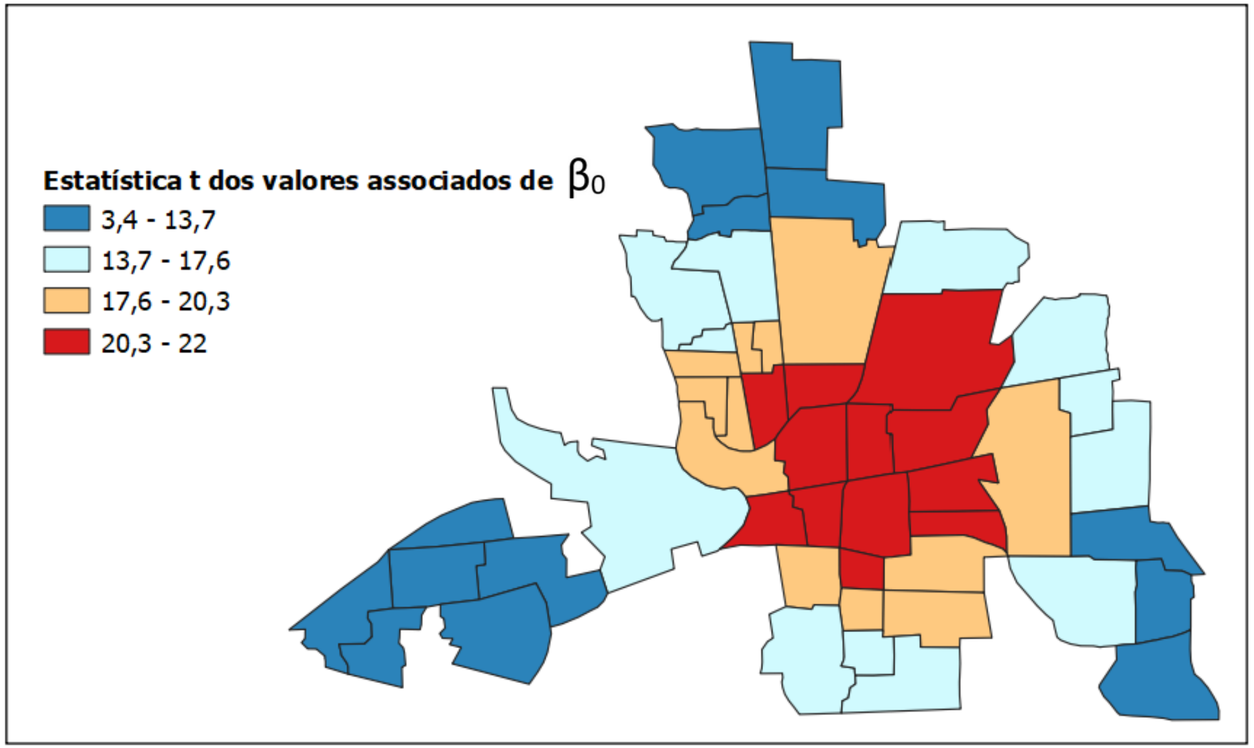
Geographically Weighted Regression (GWR)

```
  Diagnostic information
Residual sum of squares: 3957423043153280,000000
Effective number of parameters (model: trace(S)):                    9,817331
Effective number of parameters (variance: trace(S'S)):               7,873009
Degree of freedom (model: n - trace(S)):                            39,182669
Degree of freedom (residual: n - 2trace(S) + trace(S'S)):           37,238347
ML based sigma estimate:       8986864,613120
Unbiased sigma estimate:      10308869,561170
-2 log-likelihood:                1708,160885
Classic AIC:                      1729,795547
AICc:                             1736,671438
BIC/MDL:                          1750,259994
CV:                     167170079121279,000000
R square:                            0,705510
Adjusted R square:                   0,609929

****************************************************************
```

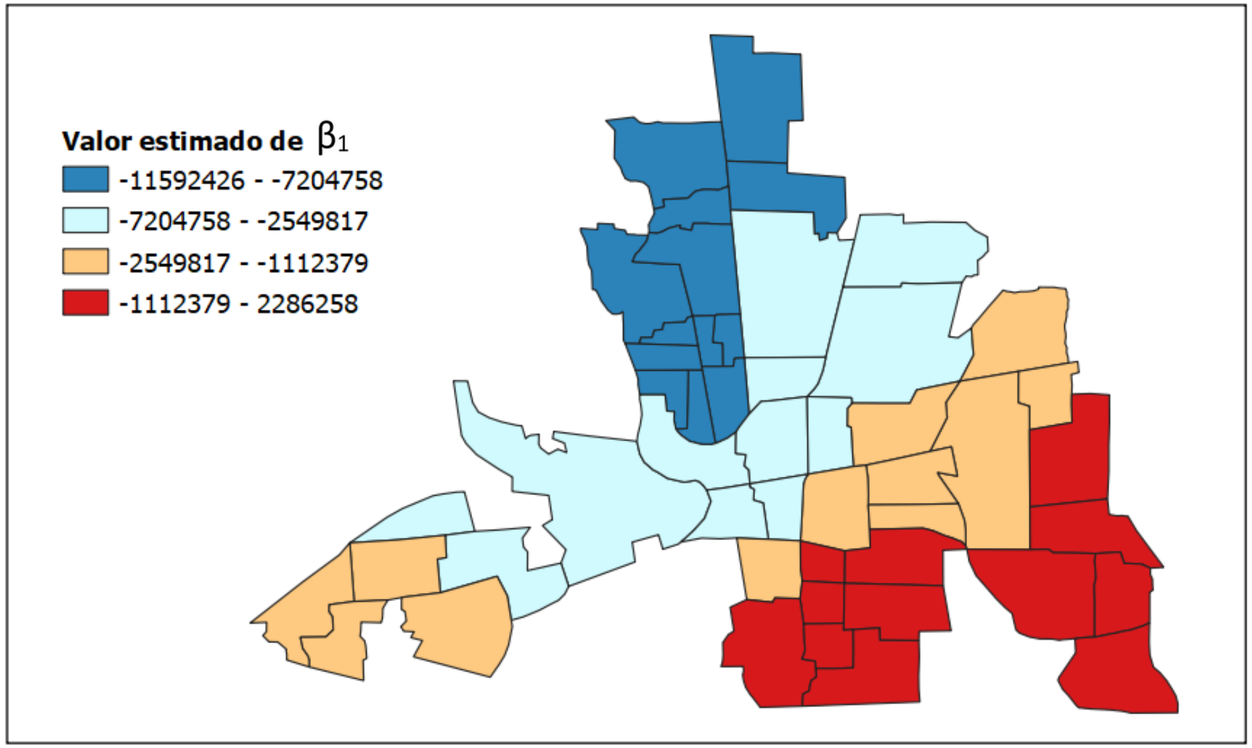
```
*******************************************************************************
 GWR ANOVA Table
*******************************************************************************
Source                          SS          DF              MS           F
----------------- ------------------- ---------- --------------- ----------
Global Residuals  6014892735784370,000     46,000
GWR Improvement   2057469692631090,000      8,762 234826645855322,000
GWR Residuals     3957423043153280,000     37,238 106272791629218,000
2,209659

*******************************************************************************
```

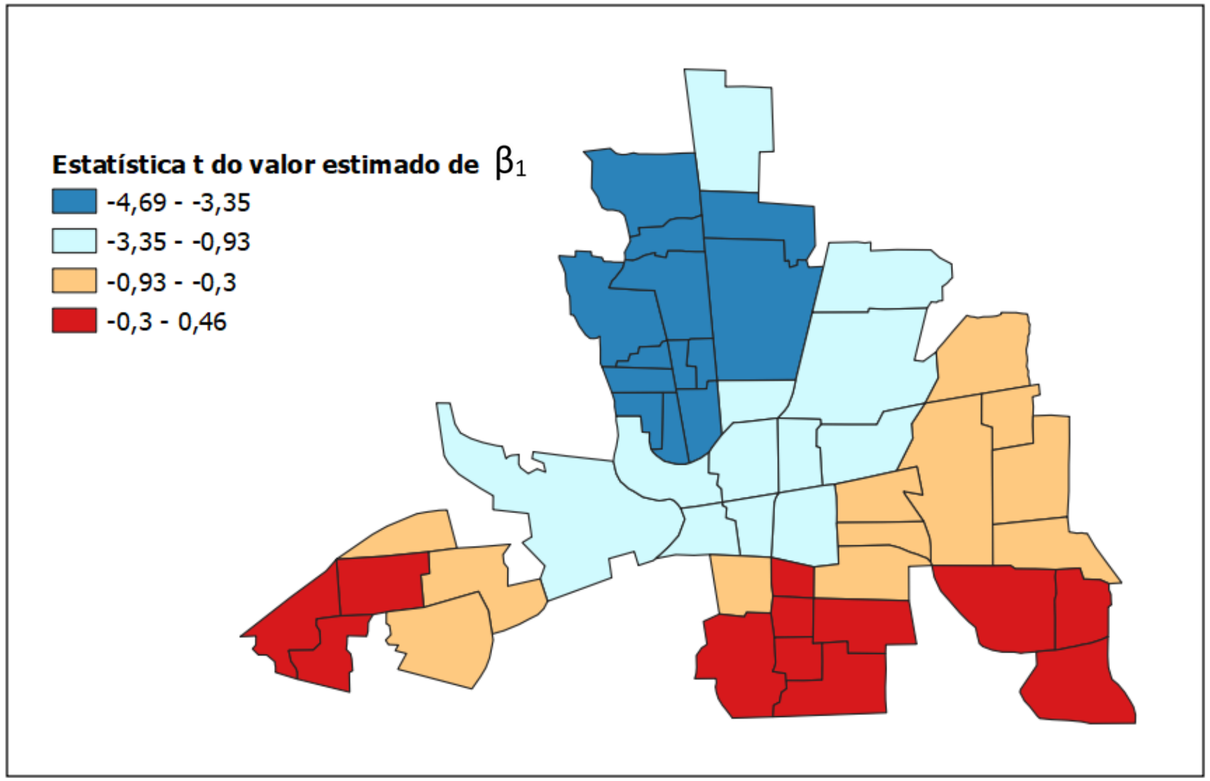
Geographically Weighted Regression (GWR)



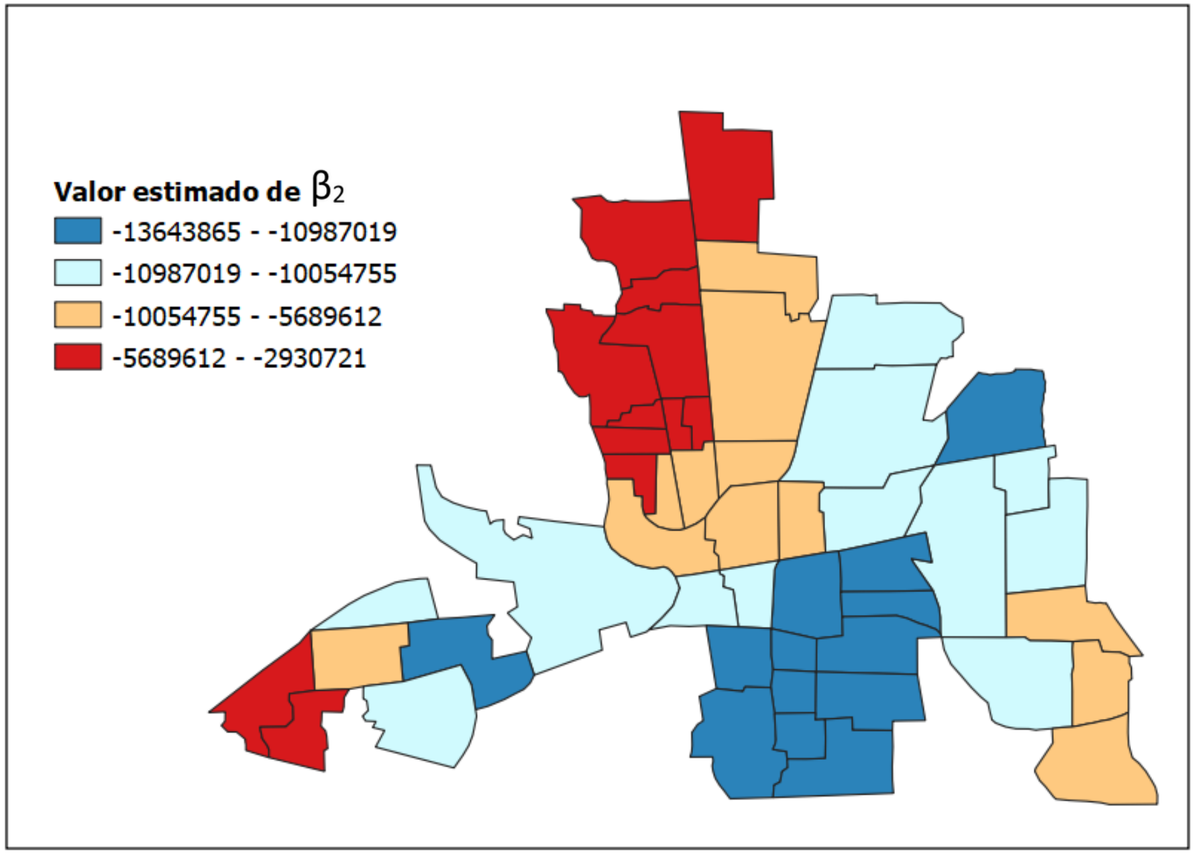


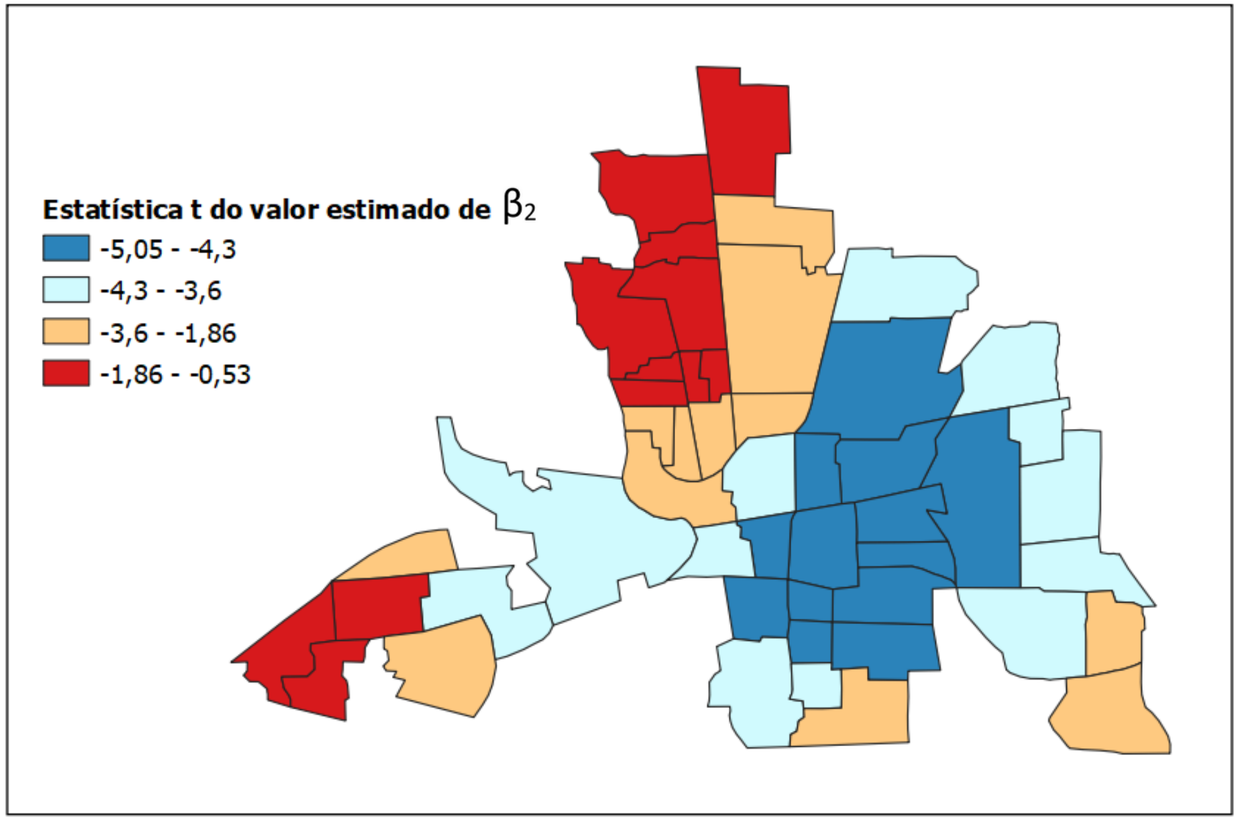
Geographically Weighted Regression (GWR)





Geographically Weighted Regression (GWR)





OBRIGADO(A)!