*Spatial Statistics Lab 8*

**Onyedikachi J Okeke**

2024-03-12

### 0.0 To Load the library

library(sp)  
library(spatstat)  
library(sf)  
library(spatstat.geom)  
library(ctv)  
library(terra)  
library(spdep)  
library(rgdal)  
library(rgdal)  
library(terra)  
library(RColorBrewer)  
library(classInt)  
library(epitools)  
library(DCluster)  
library(lmtest)  
library(car)  
library(spatialreg)  
library(spdep)  
library(classInt)  
library(spgwr)

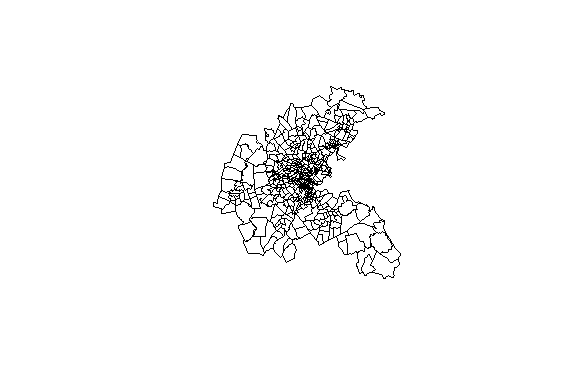
### 0.1 Loading Boston data

Boston <- readOGR(system.file("shapes/boston\_tracts.shp", package="spData")[1])  
class(Boston)  
plot(Boston)

##### Result

## OGR data source with driver: ESRI Shapefile   
## Source: "C:\Users\GIS\AppData\Local\R\win-library\4.3\spData\shapes\boston\_tracts.shp", layer: "boston\_tracts"  
## with 506 features  
## It has 36 fields

## [1] "SpatialPolygonsDataFrame"  
## attr(,"package")  
## [1] "sp"



## Part One

### 0.1 Row-standardized data and OLS for Housing

# creates a row-standardized "listw" object (default style = "W")  
Boston\_queen\_nb <- poly2nb(Boston, queen=TRUE)  
Boston\_w <- nb2listw(Boston\_queen\_nb)  
  
# creates a non row-standardized "listw" object  
Boston\_b <- nb2listw(Boston\_queen\_nb,style="B")  
  
# displays the "data.frame" attribute of the shape file  
  
boston\_data <- attr(Boston, "data")  
# Extracting the first 10 rows  
first\_5 <- head(boston\_data, 5)  
# Extracting the last 10 rows  
last\_5 <- tail(boston\_data, 5)  
# Combining the first 10 and last 10 rows  
combined\_rows <- rbind(first\_5, last\_5)  
  
# Display the combined rows  
print(combined\_rows)  
  
# using log of median housing value instead of housing value  
Boston$LOGMEDV<-log(Boston$CMEDV)  
  
# OLS for housing value  
Boston\_OLS <- lm(LOGMEDV~CRIM + CHAS + NOX + RM + AGE + DIS + B + LSTAT, data=Boston)  
summary(Boston\_OLS)  
# saving residuals and fitted values  
Boston$olsresid<-residuals(Boston\_OLS)  
Boston$ols\_fitted <- fitted(Boston\_OLS)  
# Moran test for residuals  
lm.morantest(Boston\_OLS, Boston\_w)

##### Result

## poltract TOWN TOWNNO TRACT LON LAT MEDV CMEDV  
## 0 0001 Boston Allston-Brighton 74 1 -71.0830 42.2172 17.8 17.8  
## 1 0002 Boston Allston-Brighton 74 2 -71.0950 42.2120 21.7 21.7  
## 2 0003 Boston Allston-Brighton 74 3 -71.1007 42.2100 22.7 22.7  
## 3 0004 Boston Allston-Brighton 74 4 -71.0930 42.2070 22.6 22.6  
## 4 0005 Boston Allston-Brighton 74 5 -71.0905 42.2033 25.0 25.0  
## 5 0006 Boston Allston-Brighton 74 6 -71.0865 42.2100 19.9 19.9  
## 6 0007 Boston Allston-Brighton 74 7 -71.0810 42.2080 20.8 20.8  
## 7 0008 Boston Allston-Brighton 74 8 -71.0865 42.2150 16.8 16.8  
## 8 0101 Boston Back Bay 75 101 -71.0590 42.2098 21.9 21.9  
## 9 0102 Boston Back Bay 75 102 -71.0595 42.2075 27.5 27.5  
## 496 5022 Rockland 67 5022 -70.9501 42.0825 17.2 17.2  
## 497 5031 Hanover 68 5031 -70.9275 42.0795 23.1 23.1  
## 498 5041 Norwell 69 5041 -70.9200 42.1016 24.5 24.5  
## 499 5051 Scituate 70 5051 -70.8550 42.1300 26.6 26.6  
## 500 5052 Scituate 70 5052 -70.8330 42.1150 22.9 22.9  
## 501 5061 Marshfield 71 5061 -70.8300 42.0775 24.1 24.1  
## 502 5062 Marshfield 71 5062 -70.8100 42.0590 18.6 18.6  
## 503 5071 Duxbury 72 5071 -70.8300 42.0485 30.1 30.1  
## 504 5081 Pembroke 73 5081 -70.8530 42.0520 18.2 18.2  
## 505 5082 Pembroke 73 5082 -70.8525 42.0300 20.6 20.6  
## CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT  
## 0 8.98296 0 18.10 1 0.770 6.212 97.4 2.1222 24 666 20.2 377.73 17.60  
## 1 3.84970 0 18.10 1 0.770 6.395 91.0 2.5052 24 666 20.2 391.34 13.27  
## 2 5.20177 0 18.10 1 0.770 6.127 83.4 2.7227 24 666 20.2 395.43 11.48  
## 3 4.26131 0 18.10 0 0.770 6.112 81.3 2.5091 24 666 20.2 390.74 12.67  
## 4 4.54192 0 18.10 0 0.770 6.398 88.0 2.5182 24 666 20.2 374.56 7.79  
## 5 3.83684 0 18.10 0 0.770 6.251 91.1 2.2955 24 666 20.2 350.65 14.19  
## 6 3.67822 0 18.10 0 0.770 5.362 96.2 2.1036 24 666 20.2 380.79 10.19  
## 7 4.22239 0 18.10 1 0.770 5.803 89.0 1.9047 24 666 20.2 353.04 14.64  
## 8 3.47428 0 18.10 1 0.718 8.780 82.9 1.9047 24 666 20.2 354.55 5.29  
## 9 4.55587 0 18.10 0 0.718 3.561 87.9 1.6132 24 666 20.2 354.70 7.12  
## 496 0.06162 0 4.39 0 0.442 5.898 52.3 8.0136 3 352 18.8 364.61 12.67  
## 497 0.01870 85 4.15 0 0.429 6.516 27.7 8.5353 4 351 17.9 392.43 6.36  
## 498 0.01501 80 2.01 0 0.435 6.635 29.7 8.3440 4 280 17.0 390.94 5.99  
## 499 0.02899 40 1.25 0 0.429 6.939 34.5 8.7921 1 335 19.7 389.85 5.89  
## 500 0.06211 40 1.25 0 0.429 6.490 44.4 8.7921 1 335 19.7 396.90 5.98  
## 501 0.07950 60 1.69 0 0.411 6.579 35.9 10.7103 4 411 18.3 370.78 5.49  
## 502 0.07244 60 1.69 0 0.411 5.884 18.5 10.7103 4 411 18.3 392.33 7.79  
## 503 0.01709 90 2.02 0 0.410 6.728 36.1 12.1265 5 187 17.0 384.46 4.50  
## 504 0.04301 80 1.91 0 0.413 5.663 21.9 10.5857 4 334 22.0 382.80 8.05  
## 505 0.10659 80 1.91 0 0.413 5.936 19.5 10.5857 4 334 22.0 376.04 5.57  
## units cu5k c5\_7\_5 C7\_5\_10 C10\_15 C15\_20 C20\_25 C25\_35 C35\_50 co50k median  
## 0 126 3 3 4 26 43 29 16 1 1 17800  
## 1 399 4 10 7 37 95 139 93 9 5 21700  
## 2 368 3 1 2 25 84 127 102 24 0 22700  
## 3 220 3 2 2 23 45 67 63 12 3 22600  
## 4 44 0 0 1 1 11 9 12 9 1 25000  
## 5 221 2 3 7 31 69 72 30 6 1 19900  
## 6 39 0 0 0 4 14 9 11 0 1 20800  
## 7 203 2 6 14 45 73 41 17 5 0 16800  
## 8 33 0 0 4 5 6 4 5 4 5 21900  
## 9 5 0 0 0 1 0 1 2 1 0 27500  
## 496 1043 7 7 39 267 455 199 66 2 1 17200  
## 497 2171 4 11 18 215 495 557 662 190 19 23100  
## 498 1698 3 6 19 98 330 437 432 283 90 24500  
## 499 2489 4 12 36 202 445 419 778 461 132 26600  
## 500 1164 3 5 34 131 253 271 280 141 46 22900  
## 501 1545 11 16 31 178 295 291 412 253 58 24100  
## 502 1788 2 18 75 397 546 449 258 36 7 18600  
## 503 1891 10 10 29 123 207 315 497 394 306 30100  
## 504 1792 5 35 82 376 608 422 227 30 7 18200  
## 505 716 0 2 8 64 258 217 143 22 2 20600  
## BB censored NOX\_ID POP  
## 0 0.8 no 1 3962  
## 1 1.4 no 1 9245  
## 2 0.3 no 1 6842  
## 3 0.8 no 1 8342  
## 4 1.8 no 1 7836  
## 5 3.7 no 1 9276  
## 6 1.2 no 1 9730  
## 7 3.6 no 1 8441  
## 8 3.5 no 2 10244  
## 9 3.4 no 2 7812  
## 496 2.6 no 91 5274  
## 497 0.6 no 92 10107  
## 498 0.6 no 93 7796  
## 499 0.4 no 92 11649  
## 500 0.2 no 92 5324  
## 501 1.6 no 94 7087  
## 502 0.3 no 94 8136  
## 503 1.1 no 95 7636  
## 504 0.5 no 96 7975  
## 505 1.1 no 96 3218

##   
## Call:  
## lm(formula = LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B +   
## LSTAT, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.74155 -0.12040 -0.02555 0.10726 0.96911   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.9243675 0.1666552 17.547 < 2e-16 \*\*\*  
## CRIM -0.0094454 0.0012515 -7.547 2.14e-13 \*\*\*  
## CHAS1 0.1444594 0.0364462 3.964 8.46e-05 \*\*\*  
## NOX -0.5353117 0.1401451 -3.820 0.000151 \*\*\*  
## RM 0.1351813 0.0168870 8.005 8.51e-15 \*\*\*  
## AGE -0.0005123 0.0005570 -0.920 0.358217   
## DIS -0.0396327 0.0074885 -5.292 1.81e-07 \*\*\*  
## B 0.0004169 0.0001134 3.677 0.000262 \*\*\*  
## LSTAT -0.0301493 0.0021561 -13.983 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2035 on 497 degrees of freedom  
## Multiple R-squared: 0.7554, Adjusted R-squared: 0.7515   
## F-statistic: 191.9 on 8 and 497 DF, p-value: < 2.2e-16

##   
## Global Moran I for regression residuals  
##   
## data:   
## model: lm(formula = LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B +  
## LSTAT, data = Boston)  
## weights: Boston\_w  
##   
## Moran I statistic standard deviate = 18.952, p-value < 2.2e-16  
## alternative hypothesis: greater  
## sample estimates:  
## Observed Moran I Expectation Variance   
## 0.490989041 -0.011131307 0.000701953

**Q1. Describe the OLS regression model results and their significance.**

The results of the OLS regression model indicate that several variables have a statistically significant impact on the median value of owner-occupied homes (LOGMEDV) in the Boston area.

The variable CRIM (per capita crime rate) has a negative coefficient (-0.0094454) and is highly significant (p-value < 2.2e-16), suggesting that higher crime rates lead to lower median home values. The dummy variable CHAS1 (whether the home is adjacent to the Charles River) has a positive coefficient (0.1444594) and is significant (p-value = 8.46e-05), indicating that homes near the Charles River tend to have higher median values.

The variable NOX (nitrogen oxides concentration) has a negative coefficient (-0.5353117) and is significant (p-value = 0.000151), implying that higher levels of air pollution are associated with lower median home values. The variable RM (average number of rooms per dwelling) has a positive coefficient (0.1351813) and is highly significant (p-value = 8.51e-15), suggesting that homes with more rooms tend to have higher median values.

The variable DIS (weighted distances to five Boston employment centers) has a negative coefficient (-0.0396327) and is significant (p-value = 1.81e-07), indicating that homes farther away from employment centers tend to have lower median values. The variable B (proportion of black residents) has a positive coefficient (0.0004169) and is significant (p-value = 0.000262), suggesting that a higher proportion of black residents is associated with higher median home values.

Finally, the variable LSTAT (percentage of lower status population) has a negative coefficient (-0.0301493) and is highly significant (p-value < 2.2e-16), implying that a higher percentage of lower status population leads to lower median home values.

The model has an adjusted R-squared of 0.7515, indicating that it explains approximately 75% of the variation in median home values. The Global Moran's I statistic for the regression residuals is 18.952 with a p-value < 2.2e-16, suggesting the presence of spatial autocorrelation in the residuals.

### Lagrange multiplier tests for residuals

lm.LMtests(Boston\_OLS, Boston\_w, test="all")  
#Breusch Pagan test for heteroskedasticity  
bptest(Boston\_OLS)

##### Result

##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B +  
## LSTAT, data = Boston)  
## weights: Boston\_w  
##   
## LMerr = 327.87, df = 1, p-value < 2.2e-16  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B +  
## LSTAT, data = Boston)  
## weights: Boston\_w  
##   
## LMlag = 260.91, df = 1, p-value < 2.2e-16  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B +  
## LSTAT, data = Boston)  
## weights: Boston\_w  
##   
## RLMerr = 99.454, df = 1, p-value < 2.2e-16  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B +  
## LSTAT, data = Boston)  
## weights: Boston\_w  
##   
## RLMlag = 32.498, df = 1, p-value = 1.193e-08  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B +  
## LSTAT, data = Boston)  
## weights: Boston\_w  
##   
## SARMA = 360.37, df = 2, p-value < 2.2e-16

##   
## studentized Breusch-Pagan test  
##   
## data: Boston\_OLS  
## BP = 58.379, df = 8, p-value = 9.683e-10

**Q2. Describe the diagnostics test results for spatial autocorrelation.**

The diagnostics test results indicate the presence of spatial autocorrelation in the OLS regression model for the Boston housing data.

The Lagrange Multiplier (LM) test statistics are used to test for spatial dependence in the regression residuals. The results show:

1. LMerr = 327.87 (p-value < 2.2e-16): This highly significant value suggests the presence of spatial error autocorrelation, meaning that the residuals are spatially correlated.
2. LMlag = 260.91 (p-value < 2.2e-16): This highly significant value indicates the presence of spatial lag autocorrelation, which means that the dependent variable (LOGMEDV) is spatially correlated.
3. RLMerr = 99.454 (p-value < 2.2e-16): This is the robust version of the LMerr statistic, which also suggests significant spatial error autocorrelation.
4. RLMlag = 32.498 (p-value = 1.193e-08): This is the robust version of the LMlag statistic, which also indicates significant spatial lag autocorrelation.
5. SARMA = 360.37 (p-value < 2.2e-16): The SARMA statistic tests for the presence of both spatial error and spatial lag autocorrelation. The highly significant value suggests the presence of both types of spatial autocorrelation.

The studentized Breusch-Pagan test result (BP = 58.379, p-value = 9.683e-10) also indicates the presence of heteroskedasticity in the regression residuals.

Overall, these diagnostic test results strongly suggest that the OLS regression model violates the assumptions of no spatial autocorrelation and homoskedasticity. The presence of spatial autocorrelation and heteroskedasticity can lead to inefficient parameter estimates and invalid statistical inferences. Therefore, it may be appropriate to consider spatial regression models, such as the Spatial Error Model (SEM) or Spatial Lag Model (SLM), to account for the spatial dependence in the data.

### Spatial lag Regression

Boston\_spatial\_lag <- lagsarlm(LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B + LSTAT, data=Boston, Boston\_w)  
summary(Boston\_spatial\_lag)  
# saving residuals and fitted values  
Boston$lagresid<-residuals(Boston\_spatial\_lag)  
Boston$lag\_fitted <- fitted(Boston\_spatial\_lag)

##### Result

##   
## Call:lagsarlm(formula = LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS +   
## B + LSTAT, data = Boston, listw = Boston\_w)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.602982 -0.087559 -0.014017 0.079602 0.817612   
##   
## Type: lag   
## Coefficients: (asymptotic standard errors)   
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 9.1935e-01 1.6647e-01 5.5225 3.342e-08  
## CRIM -5.1768e-03 9.9894e-04 -5.1823 2.191e-07  
## CHAS1 4.8343e-02 2.8108e-02 1.7199 0.085452  
## NOX -1.5241e-01 1.0914e-01 -1.3965 0.162567  
## RM 1.2963e-01 1.3040e-02 9.9405 < 2.2e-16  
## AGE -4.6463e-04 4.2565e-04 -1.0916 0.275014  
## DIS -2.9370e-02 5.8160e-03 -5.0499 4.421e-07  
## B 2.6460e-04 8.7251e-05 3.0326 0.002424  
## LSTAT -1.6652e-02 1.7990e-03 -9.2560 < 2.2e-16  
##   
## Rho: 0.54543, LR test value: 233.72, p-value: < 2.22e-16  
## Asymptotic standard error: 0.03174  
## z-value: 17.185, p-value: < 2.22e-16  
## Wald statistic: 295.31, p-value: < 2.22e-16  
##   
## Log likelihood: 208.9408 for lag model  
## ML residual variance (sigma squared): 0.024053, (sigma: 0.15509)  
## Number of observations: 506   
## Number of parameters estimated: 11   
## AIC: -395.88, (AIC for lm: -164.16)  
## LM test for residual autocorrelation  
## test value: 29.36, p-value: 6.0116e-08

## This method assumes the response is known - see manual page

**Q3. Describe the spatial lag regression model results and its significance.**

The spatial lag coefficient (Rho) is 0.54543, which is highly significant (p-value < 2.22e-16) based on the likelihood ratio (LR) test, Wald test, and z-value. This positive and significant value suggests the presence of spatial lag dependence, meaning that the median home values in a given location are influenced by the median home values in neighboring locations.

Looking at the individual coefficients:

1. CRIM (per capita crime rate) has a negative coefficient (-0.0051768) and is significant (p-value = 2.191e-07), indicating that higher crime rates are associated with lower median home values, even after accounting for spatial dependence.
2. CHAS1 (whether the home is adjacent to the Charles River) has a positive coefficient (0.048343) but is not significant at the 5% level (p-value = 0.085452).
3. NOX (nitrogen oxides concentration) has a negative coefficient (-0.15241) but is not significant (p-value = 0.162567).
4. RM (average number of rooms per dwelling) has a positive coefficient (0.12963) and is highly significant (p-value < 2.2e-16), indicating that homes with more rooms tend to have higher median values, even after accounting for spatial dependence.
5. AGE (proportion of owner-occupied units built before 1940) has a negative coefficient (-0.0004646) but is not significant (p-value = 0.275014).
6. DIS (weighted distances to five Boston employment centers) has a negative coefficient (-0.029370) and is significant (p-value = 4.421e-07), suggesting that homes farther away from employment centers tend to have lower median values.
7. B (proportion of black residents) has a positive coefficient (0.0002646) and is significant (p-value = 0.002424), implying that a higher proportion of black residents is associated with higher median home values.
8. LSTAT (percentage of lower status population) has a negative coefficient (-0.016652) and is highly significant (p-value < 2.2e-16), indicating that a higher percentage of lower status population leads to lower median home values.

The model has a log-likelihood of 208.9408, and the AIC (-395.88) is lower than the AIC for the OLS model (-164.16), suggesting that the spatial lag model provides a better fit to the data.

However, the LM test for residual autocorrelation is still significant (test value: 29.36, p-value: 6.0116e-08), indicating that spatial dependence may still exist in the residuals, and further model refinements or alternative spatial models may be required.

### Error regression

Boston\_spatial\_error <- errorsarlm(LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B + LSTAT, data=Boston, Boston\_w)  
summary(Boston\_spatial\_error)  
# saving residuals and fitted values  
Boston$error\_resid<-residuals(Boston\_spatial\_error)  
Boston$error\_fitted <- fitted(Boston\_spatial\_error)

##### Result

##   
## Call:errorsarlm(formula = LOGMEDV ~ CRIM + CHAS + NOX + RM + AGE +   
## DIS + B + LSTAT, data = Boston, listw = Boston\_w)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.7127664 -0.0744056 -0.0083521 0.0677584 0.7394041   
##   
## Type: error   
## Coefficients: (asymptotic standard errors)   
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 2.80261293 0.18191754 15.4060 < 2.2e-16  
## CRIM -0.00707212 0.00105312 -6.7154 1.875e-11  
## CHAS1 -0.01442505 0.03674072 -0.3926 0.6946019  
## NOX -0.80066014 0.18208348 -4.3972 1.096e-05  
## RM 0.16312717 0.01431568 11.3950 < 2.2e-16  
## AGE -0.00149580 0.00052020 -2.8754 0.0040351  
## DIS -0.04065136 0.01272359 -3.1950 0.0013985  
## B 0.00044147 0.00012811 3.4461 0.0005688  
## LSTAT -0.01907178 0.00198011 -9.6317 < 2.2e-16  
##   
## Lambda: 0.78429, LR test value: 277.61, p-value: < 2.22e-16  
## Asymptotic standard error: 0.032101  
## z-value: 24.432, p-value: < 2.22e-16  
## Wald statistic: 596.92, p-value: < 2.22e-16  
##   
## Log likelihood: 230.8859 for error model  
## ML residual variance (sigma squared): 0.020082, (sigma: 0.14171)  
## Number of observations: 506   
## Number of parameters estimated: 11   
## AIC: -439.77, (AIC for lm: -164.16)

## This method assumes the response is known - see manual page

Q4. Describe the spatial error regression model results and its significance.

## Part Two

### create a data frame

# create a data frame  
boston.df <- data.frame(Boston)  
names(boston.df) # gives us the column names  
colnames(boston.df) # alternative call; same thing  
  
attach(boston.df) #Now you can call the variable name without using “$”  
  
LOGCMEDV <- log(CMEDV)  
  
# we create this location using the coordinates of the centroids:  
XYgridtable <- cbind(LON, LAT)  
#using gwr.sel to find optimal parameters  
adaptive.bw<-gwr.sel(LOGCMEDV~CRIM+CHAS+NOX+RM+AGE+DIS+B+LSTAT,adapt=TRUE, method="cv", coords=XYgridtable)  
###gwr.Gauss() default  
  
#number of observations (polygon)  
dim(boston.df)[1]  
# how many points does each window take  
print(dim(boston.df)[1]\*adaptive.bw )  
#run gwr using the optimal parameter found above  
bos.gwr<-gwr(LOGCMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS + B + LSTAT, adapt = adaptive.bw, gweight = gwr.Gauss, hatmatrix = TRUE, coords = XYgridtable)  
print(bos.gwr)  
#Plot local R2  
classes\_fx <- classIntervals(bos.gwr$SDF$localR2, n=5, style="quantile")  
cols <- findColours(classes\_fx,pal)  
pal <- brewer.pal(5, "YlOrRd")  
plot(Boston,col=cols, border="grey")  
legend(x="topleft", cex=0.65, fill=attr(cols,"palette"), bty="n", legend=names(attr(cols, "table")), title="Local R^2",ncol=1)  
#Plot Residual  
res <- bos.gwr$SDF$gwr.e  
classes\_fx <- classIntervals(res, n=5, style="quantile")  
pal <- brewer.pal(5, "YlOrRd")  
cols <- findColours(classes\_fx,pal)  
plot(Boston,col=cols, border="transparent")  
legend(x="topleft", cex=0.65, fill=attr(cols, "palette"), bty="n", legend=names(attr(cols, "table")), title="Residual", ncol=1)

##### Result

## [1] "poltract" "TOWN" "TOWNNO" "TRACT" "LON"   
## [6] "LAT" "MEDV" "CMEDV" "CRIM" "ZN"   
## [11] "INDUS" "CHAS" "NOX" "RM" "AGE"   
## [16] "DIS" "RAD" "TAX" "PTRATIO" "B"   
## [21] "LSTAT" "units" "cu5k" "c5\_7\_5" "C7\_5\_10"   
## [26] "C10\_15" "C15\_20" "C20\_25" "C25\_35" "C35\_50"   
## [31] "co50k" "median" "BB" "censored" "NOX\_ID"   
## [36] "POP" "LOGMEDV" "olsresid" "ols\_fitted" "lagresid"   
## [41] "lag\_fitted" "error\_resid" "error\_fitted"

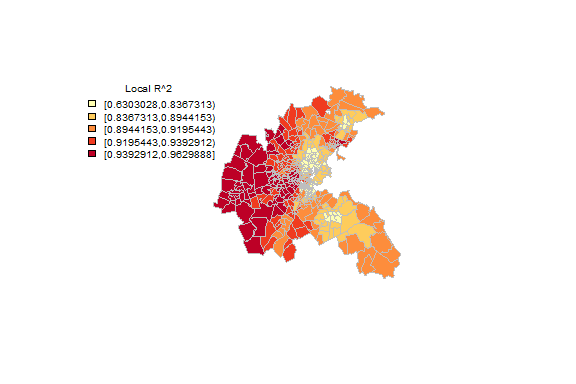
## [1] "poltract" "TOWN" "TOWNNO" "TRACT" "LON"   
## [6] "LAT" "MEDV" "CMEDV" "CRIM" "ZN"   
## [11] "INDUS" "CHAS" "NOX" "RM" "AGE"   
## [16] "DIS" "RAD" "TAX" "PTRATIO" "B"   
## [21] "LSTAT" "units" "cu5k" "c5\_7\_5" "C7\_5\_10"   
## [26] "C10\_15" "C15\_20" "C20\_25" "C25\_35" "C35\_50"   
## [31] "co50k" "median" "BB" "censored" "NOX\_ID"   
## [36] "POP" "LOGMEDV" "olsresid" "ols\_fitted" "lagresid"   
## [41] "lag\_fitted" "error\_resid" "error\_fitted"

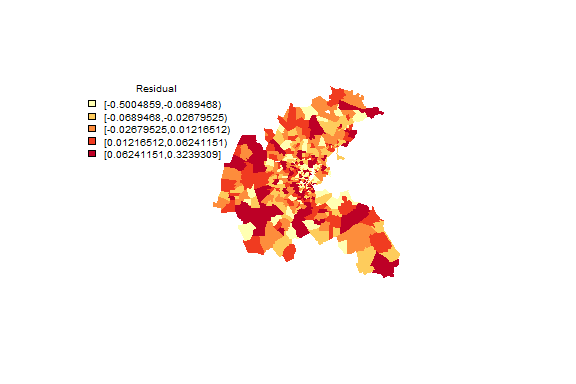
## Adaptive q: 0.381966 CV score: 19.45164   
## Adaptive q: 0.618034 CV score: 20.59138   
## Adaptive q: 0.236068 CV score: 18.32478   
## Adaptive q: 0.145898 CV score: 17.00162   
## Adaptive q: 0.09016994 CV score: 15.20693   
## Adaptive q: 0.05572809 CV score: 13.43434   
## Adaptive q: 0.03444185 CV score: 12.16576   
## Adaptive q: 0.02128624 CV score: 11.48963   
## Adaptive q: 0.01315562 CV score: 12.34929   
## Adaptive q: 0.02438278 CV score: 11.59113   
## Adaptive q: 0.01818062 CV score: 11.69557   
## Adaptive q: 0.02180867 CV score: 11.51344   
## Adaptive q: 0.02009999 CV score: 11.45838   
## Adaptive q: 0.01952185 CV score: 11.48955   
## Adaptive q: 0.02040351 CV score: 11.46344   
## Adaptive q: 0.02014758 CV score: 11.45902   
## Adaptive q: 0.01987916 CV score: 11.45627   
## Adaptive q: 0.01964877 CV score: 11.4714   
## Adaptive q: 0.01979116 CV score: 11.45583   
## Adaptive q: 0.01973677 CV score: 11.45926   
## Adaptive q: 0.01983185 CV score: 11.456   
## Adaptive q: 0.01979116 CV score: 11.45583

## [1] 506

## [1] 10.01433

## Call:  
## gwr(formula = LOGCMEDV ~ CRIM + CHAS + NOX + RM + AGE + DIS +   
## B + LSTAT, coords = XYgridtable, gweight = gwr.Gauss, adapt = adaptive.bw,   
## hatmatrix = TRUE)  
## Kernel function: gwr.Gauss   
## Adaptive quantile: 0.01979116 (about 10 of 506 data points)  
## Summary of GWR coefficient estimates at data points:  
## Min. 1st Qu. Median 3rd Qu. Max.  
## X.Intercept. 0.46639045 1.74004057 2.74084330 3.53255460 8.37525856  
## CRIM -0.38756924 -0.01231664 -0.00998365 -0.00756998 0.37331527  
## CHAS1 -0.17682389 0.01366529 0.15869549 0.57823993 3.06857177  
## NOX -4.45010132 -0.92336607 -0.58516056 -0.12435515 2.29273633  
## RM -0.22730751 0.08279618 0.18060022 0.26045226 0.36282998  
## AGE -0.02174969 -0.00371533 -0.00217082 -0.00060200 0.01354966  
## DIS -1.26025807 -0.13272898 -0.05461739 -0.02672744 0.40216030  
## B -0.00195998 0.00028663 0.00057108 0.00087532 0.00272656  
## LSTAT -0.04756575 -0.03235929 -0.01979628 -0.00846942 0.00350482  
## Global  
## X.Intercept. 2.9244  
## CRIM -0.0094  
## CHAS1 0.1445  
## NOX -0.5353  
## RM 0.1352  
## AGE -0.0005  
## DIS -0.0396  
## B 0.0004  
## LSTAT -0.0301  
## Number of data points: 506   
## Effective number of parameters (residual: 2traceS - traceS'S): 160.0712   
## Effective degrees of freedom (residual: 2traceS - traceS'S): 345.9288   
## Sigma (residual: 2traceS - traceS'S): 0.1249992   
## Effective number of parameters (model: traceS): 118.8724   
## Effective degrees of freedom (model: traceS): 387.1276   
## Sigma (model: traceS): 0.1181608   
## Sigma (ML): 0.1033535   
## AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): -545.8805   
## AIC (GWR p. 96, eq. 4.22): -741.9968   
## Residual sum of squares: 5.405068   
## Quasi-global R2: 0.9357897





**Q5. Take screenshots of the plots and describe the pattern of local R2 and residual.**

In the first map showing Local  values, the range of values is categorized into five intervals, with darker shades indicating higher Local values. For example, the darkest red areas have values between 0.9393 and 0.9629, meaning the model explains between 93.93% and 96.29% of the variance in those areas. The lightest areas, in contrast, have values between 0.6303 and 0.8387, indicating a lower explanatory power of the model in those regions.

The second map of residuals also uses a five-category range. Here, darker shades (both red and orange) indicate areas where the model's predictions are less accurate (residuals are high, either positive or negative). Lighter shades indicate areas where the model's predictions are more accurate (residuals are close to zero). The darkest regions show residuals from -0.5004 to -0.0895 or from 0.0824 to 0.3239, highlighting areas where the observed values are much lower or much higher than the predicted values, respectively. The lightest regions have residuals very close to zero (between -0.0288 and 0.0122), meaning the model's predictions are quite accurate in those areas.

### Spatial lag Regression

###plot local coefficient (CRIM)  
coef <- bos.gwr$SDF$CRIM  
classes\_fx <- classIntervals(coef, n=5, style="quantile")  
cols <- findColours(classes\_fx,pal)  
plot(Boston,col=cols, border="transparent")  
legend(x="topleft", cex=0.75, fill=attr(cols, "palette"), bty="n", legend=names(attr(cols, "table")), title = "Local Coefficient Estimates (CRIM)", ncol=1)  
###plot local coefficient (NOX)  
coef <- bos.gwr$SDF$NOX  
classes\_fx <- classIntervals(coef, n=5, style="quantile")  
cols <- findColours(classes\_fx,pal)  
  
plot(Boston,col=cols, border="transparent")  
legend(x="topleft", cex=0.7, fill=attr(cols, "palette"), bty="n", legend=names(attr(cols, "table")), title="Local Coefficient Estimates (NOX)", ncol = 1)

Result

A map of a city with numbers

Description automatically generatedA map of a city with numbers and a black text

Description automatically generatedA map of a city

Description automatically generatedA map of a country with numbers

Description automatically generatedA map of a city with numbers

Description automatically generatedA map of a city

Description automatically generated

**Q6. Plot and describe the patterns. Also plot other independent variables in the gwr**

1. CRIM (Crime Rate): Coefficient estimates range from about -0.39 to 0.37, with negative values indicating areas where higher crime rates are associated with lower property values, and positive values suggesting the opposite. Darker red areas are where crime rates have a more significant negative impact on property values.

2. NOX (Nitric Oxides Concentration): Estimates range from -1.45 to 2.29. The map likely indicates a negative relationship between NOX levels and property values in areas with darker reds. Conversely, where coefficients are positive (lighter areas), an increase in NOX levels might be associated with higher property values.

3. RM (Average Number of Rooms per Dwelling): The values here span from -0.22 to 0.36. Darker red areas might be where additional rooms have a higher positive impact on property values, indicating a preference for larger homes.

4. AGE (Age of Dwelling): Estimates vary from -0.021 to 0.013. In areas with darker reds, older properties could be negatively impacting property values, whereas lighter areas might show either a positive impact or a negligible one.

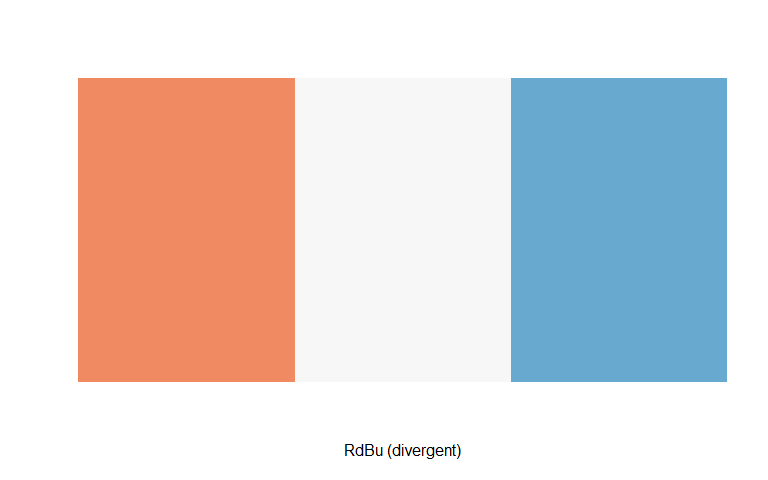
5. DIS (Distances to Employment Centers): Coefficient estimates range from -1.26 to 0.40. The negative values in darker areas suggest that longer distances from employment centers are associated with lower property values. In contrast, the positive values in lighter areas indicate that proximity to employment centers could be valued more.

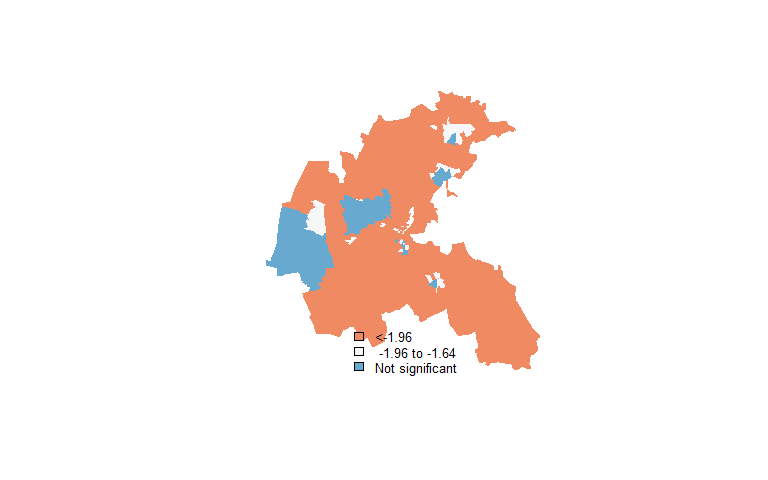
6. B (Proportion of Blacks by Town): Estimates go from -0.0019 to 0.0027. This map shows minimal variation in the coefficient values, indicating the impact of this variable might be relatively uniform, or changes might not have a substantial effect on property values.

### plot t values (CRIME)

#plot t values (CRIME)  
bos.gwr$SDF$CRIM\_t <- bos.gwr$SDF$CRIM/bos.gwr$SDF$CRIM\_se  
tCRIM<-bos.gwr$SDF$CRIM\_t  
  
display.brewer.pal(3, "RdBu")  
colors <- brewer.pal(3, "RdBu") # Stores colors in object color  
color.category.reg <- findInterval(tCRIM,  
c(min(tCRIM)-.0001, -1.96, -1.64,  
max(tCRIM)+.0001),all.inside=TRUE)  
  
#classes\_fx <- classIntervals(tCRIM, n=5, style="quantile")  
#cols <- findColours(classes\_fx,pal)  
  
plot(Boston,col=colors[color.category.reg], border="transparent")  
labels <- c("<-1.96", " -1.96 to -1.64", "Not significant")  
legend("bottom", legend=labels, fill=colors, cex=0.8,  
y.intersp = 0.99, bty="n")

##### Result





**Q7. Take a screen shot of the plot and describe the pattern.**

Blue areas show where the variable has a strongly significant negative effect, with t-values below -1.96. White areas signify moderately significant negative effects with t-values ranging from -1.96 to -1.64. The predominant orange regions represent areas where the variable's influence is not statistically significant; the effects there do not stand out from zero in a meaningful way. The concentration of blue and white areas could indicate localized factors influencing the variable, while the widespread non-significance suggests a general lack of influence across most regions.

**Q8. Describe the GWR model results and their significance. How does it compare with the regression results in part 1?**

The GWR model's results showcase spatially varied relationships between predictors and the dependent variable, with certain predictors having significant localized impacts in specific areas. If we look at the visualizations associated with the GWR model, these would reflect significant t-values in some regions (indicated by blue and white areas) and non-significant effects (indicated by orange areas). The significance of the predictors in blue areas with t-values less than -1.96 implies a strong negative relationship, while white areas with t-values between -1.96 and -1.64 indicate a moderate negative relationship.

Comparing this to the previously discussed OLS regression, which indicated significant effects for predictors such as CRIM and NOX on the dependent variable (with p-values < 0.05), the GWR analysis adds a layer of spatial detail. While OLS may suggest a uniform effect of CRIM and NOX across all areas, the GWR results imply that the strength and significance of these relationships vary by location. For example, the OLS may yield a negative coefficient for NOX with a low p-value, but GWR could reveal that this negative impact is only significant in certain parts of the study area.

This suggests that while OLS captures the average effect, GWR allows for the possibility that the relationship between housing values and environmental or social factors can differ across the city—a crucial insight for place-specific policy formulation.

**Q9. How could you use these results as a policymaker who is interested in environmental or social issues (pollution or affordable housing)?**

As a policymaker interested in addressing environmental or social issues such as pollution or affordable housing, GWR model results can be instrumental for several reasons:

1. **Targeted Interventions:** The localized estimates from the GWR model can help identify specific areas where intervention may be most needed or effective. For instance, if certain areas show a strong negative impact of pollution (NOX) on housing values, environmental regulations could be focused there.
2. **Resource Allocation:** Resources for environmental cleanup or affordable housing can be allocated more efficiently. Areas identified by the GWR as significantly impacted by CRIM or NOX would be prioritized for environmental programs or community policing initiatives.
3. **Policy Evaluation**: By mapping out the variation in effects across different regions, a policymaker can evaluate where current policies may be falling short and where they are effective, allowing for a more nuanced policy adjustment.
4. **Community Engagement**: Understanding localized impacts empowers communities. Policymakers can use this information to engage with local residents about specific issues affecting their neighborhood, fostering community-driven solutions.
5. **Evidence-Based Planning:** For future developments, GWR results could guide zoning decisions or the placement of new housing to avoid areas where environmental factors negatively impact living conditions.
6. **Monitoring and Adaptation:** Over time, GWR can be applied to new data to monitor the effectiveness of policies and adapt them as needed, ensuring that policy responses remain relevant to the changing spatial dynamics of cities.