

# HW5

Haojie Liu

2025-06-08

```
library(tidyverse)
library(sf)
library(classInt)
library(tmap)
library(spdep)
library(spatialreg)
library(spgwr)
library(sp)
```

The zipped file USCovid contains county level data on COVID deaths per 100,000 of the population registered over the first year of the pandemic (roughly March 2020 – February 2021). The data in this file have all been normalized. Here are the meanings of the variables of interest:

```
covid_data <- st_read("USCovid/totals_apr20_feb21.shp")
```

Generate a quantile map of the Covid death rate and report what you see.

```
## Reading layer 'totals_apr20_feb21' from data source
##   '/Users/liuhaojie/Desktop/STATSM222/HW5/USCovid/totals_apr20_feb21.shp'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 3108 features and 15 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -2031905 ymin: -2116976 xmax: 2516374 ymax: 732352.2
## Projected CRS: US National Atlas Equal Area

covid_data <- covid_data %>%
  mutate(death_quantile = cut(cases_per_,
                               breaks = quantile(cases_per_, probs = seq(0, 1, 0.2), na.rm = TRUE),
                               include.lowest = TRUE,
                               labels = c("Q1", "Q2", "Q3", "Q4", "Q5")))

tmap_mode("plot")

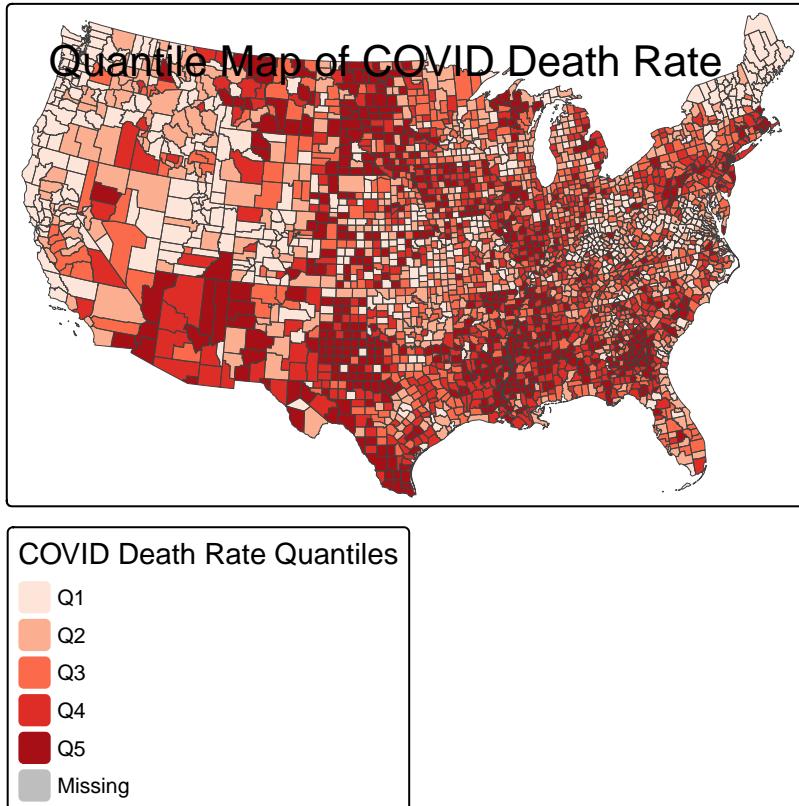
## i tmap mode set to "plot".
```

```

tm_shape(covid_data) +
  tm_fill("death_quantile", palette = "Reds", title = "COVID Death Rate Quantiles") +
  tm_borders(lwd = 0.1) +
  tm_layout(title = "Quantile Map of COVID Death Rate", legend.outside = TRUE)

##
## -- tmap v3 code detected -----
## [v3->v4] 'tm_tm_fill()': migrate the argument(s) related to the scale of the
## visual variable 'fill' namely 'palette' (rename to 'values') to fill.scale =
## tm_scale(<HERE>).[v3->v4] 'tm_fill()': migrate the argument(s) related to the legend of the
## visual variable 'fill' namely 'title' to 'fill.legend = tm_legend(<HERE>)'[v3->v4] 'tm_layout()': us
## 'cols4all::c4a_gui()' to explore them. The old palette name "Reds" is named
## "brewer.reds"Multiple palettes called "reds" found: "brewer.reds", "matplotlib.reds". The first one,

```



```

nb <- poly2nb(covid_data, queen = TRUE)

```

Is there evidence of spatial autocorrelation in the Covid death rate data? (Use a reasonable spatial weights scheme.)

```

## Warning in poly2nb(covid_data, queen = TRUE): some observations have no neighbours;
## if this seems unexpected, try increasing the snap argument.

```

```
## Warning in poly2nb(covid_data, queen = TRUE): neighbour object has 3 sub-graphs;
## if this sub-graph count seems unexpected, try increasing the snap argument.
```

```
lw <- nb2listw(nb, style = "W", zero.policy = TRUE)

moran_result <- moran.test(covid_data$cases_per_, lw, zero.policy = TRUE)
print(moran_result)
```

```
##
## Moran I test under randomisation
##
## data: covid_data$cases_per_
## weights: lw
## n reduced by no-neighbour observations
##
## Moran I statistic standard deviate = 35.882, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.3856730726    -0.0003220612     0.0001157177
```

The Moran's I test reveals strong evidence of spatial autocorrelation in county-level COVID death rates across the U.S. With a Moran's I value of 0.386 and a highly significant p-value (< 2.2e-16), the results indicate that counties with similar COVID death rates are geographically clustered rather than randomly distributed. This suggests that spatial factors—such as regional health infrastructure, demographic composition, or policy responses—may be influencing the spread and severity of COVID-19 outcomes.

```
local_moran <- localmoran(covid_data$cases_per_, lw, zero.policy = TRUE)

covid_data$local_I <- local_moran[,1]
covid_data$local_p <- local_moran[,5]

covid_data$cluster_type <- case_when(
  covid_data$local_I > 0 & covid_data$cases_per_ > mean(covid_data$cases_per_, na.rm = TRUE) & covid_da
  covid_data$local_I > 0 & covid_data$cases_per_ < mean(covid_data$cases_per_, na.rm = TRUE) & covid_da
  covid_data$local_I < 0 & covid_data$cases_per_ > mean(covid_data$cases_per_, na.rm = TRUE) & covid_da
  covid_data$local_I < 0 & covid_data$cases_per_ < mean(covid_data$cases_per_, na.rm = TRUE) & covid_da
  TRUE ~ "Not significant"
)

tmap_mode("plot")
```

Is there evidence of significant local spatial clusters of Covid death rates?

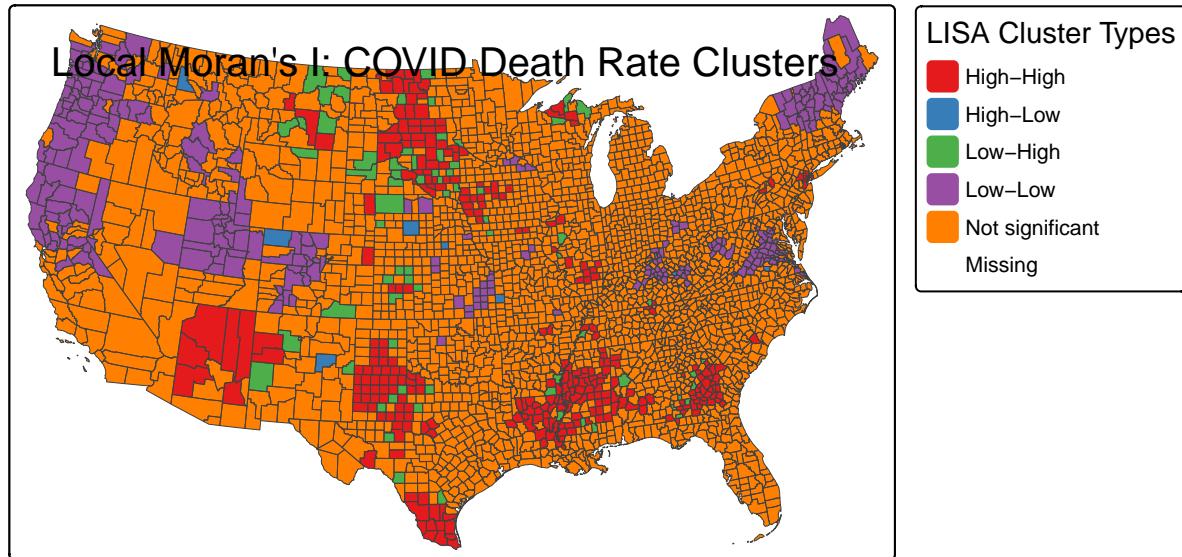
```
## i tmap mode set to "plot".

tm_shape(covid_data) +
  tm_fill("cluster_type", palette = "Set1", title = "LISA Cluster Types") +
  tm_borders(lwd = 0.1) +
  tm_layout(title = "Local Moran's I: COVID Death Rate Clusters", legend.outside = TRUE)
```

```

## 
## -- tmap v3 code detected -----
## [v3->v4] 'tm_tm_fill()': migrate the argument(s) related to the scale of the
## visual variable 'fill' namely 'palette' (rename to 'values') to fill.scale =
## tm_scale(<HERE>).[v3->v4] 'tm_fill()': migrate the argument(s) related to the legend of the
## visual variable 'fill' namely 'title' to 'fill.legend = tm_legend(<HERE>)'[v3->v4] 'tm_layout()': us
## 'cols4all::c4a_gui()' to explore them. The old palette name "Set1" is named
## "brewer.set1"

```



The Local Moran's I (LISA) analysis provides clear evidence of significant local spatial clusters in COVID-19 death rates across U.S. counties. High-high clusters, where counties with high death rates are surrounded by similarly high-rate neighbors, are prominently found in the Deep South, southern Texas, and parts of the Southwest. In contrast, low-low clusters, indicating consistently low death rates, are concentrated in regions such as the West Coast, Northeast, and Upper Midwest. Additionally, several spatial outliers (high-low and low-high counties) suggest localized deviations from surrounding patterns. These findings highlight that COVID-19 mortality was not randomly distributed, but instead shaped by strong regional dynamics and neighboring effects.

```

model <- lm(cases_per_ ~ pct_poc + pct_smoker + pct_povert + pct_obsese + pct_65plus + per_dem, data = )
summary(model)

```

Régresser la mortalité COVID contre les six variables indépendantes données ci-dessus. Montrer et interpréter votre sortie de régression (coefficients partiels de régression, statistiques d'ajustement).

```

## 
## Call:
## lm(formula = cases_per_ ~ pct_poc + pct_smoker + pct_povert +
##     pct_obsese + pct_65plus + per_dem, data = covid_data)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -2.2429 -0.5982 -0.1357  0.4312  6.5142 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 0.01362   0.01656   0.822   0.411    
## pct_poc     0.28321   0.02737  10.347  < 2e-16 *** 
## pct_smoker -0.01943   0.02273  -0.855   0.393    
## pct_povert   0.11754   0.01991   5.904 3.92e-09 *** 
## pct_obsese   0.14833   0.01955   7.586 4.34e-14 *** 
## pct_65plus   0.12800   0.01964   6.518 8.27e-11 *** 
## per_dem     -0.16805   0.02148  -7.824 6.97e-15 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 0.9227 on 3101 degrees of freedom
## Multiple R-squared:  0.1432, Adjusted R-squared:  0.1415 
## F-statistic: 86.38 on 6 and 3101 DF,  p-value: < 2.2e-16

```

A multiple linear regression was conducted to examine the relationship between county-level COVID-19 death rates and six explanatory variables. The model was statistically significant overall ( $F(6, 3101) = 86.38$ ,  $p < 0.001$ ), with an adjusted  $R^2$  of 0.1415, indicating that approximately 14% of the variation in COVID-19 death rates is explained by the predictors.

The results suggest that counties with higher proportions of people of color, individuals in poverty, obese residents, and elderly populations tend to have higher COVID-19 death rates. In contrast, a higher share of Democratic voters is associated with lower death rates. The smoking rate did not show a significant association with mortality after controlling for other factors. These findings highlight the role of both demographic and political characteristics in shaping county-level pandemic outcomes.

```

covid_data$resid <- residuals(model)

tmap_mode("plot")

```

Capture and plot the residuals from your regression. Is there evidence of spatial autocorrelation in the residuals?

```

## i tmap mode set to "plot".

tm_shape(covid_data) +
  tm_fill("resid", palette = "-RdBu", title = "Regression Residuals") +
  tm_borders(lwd = 0.1) +
  tm_layout(title = "Spatial Distribution of Residuals", legend.outside = TRUE)

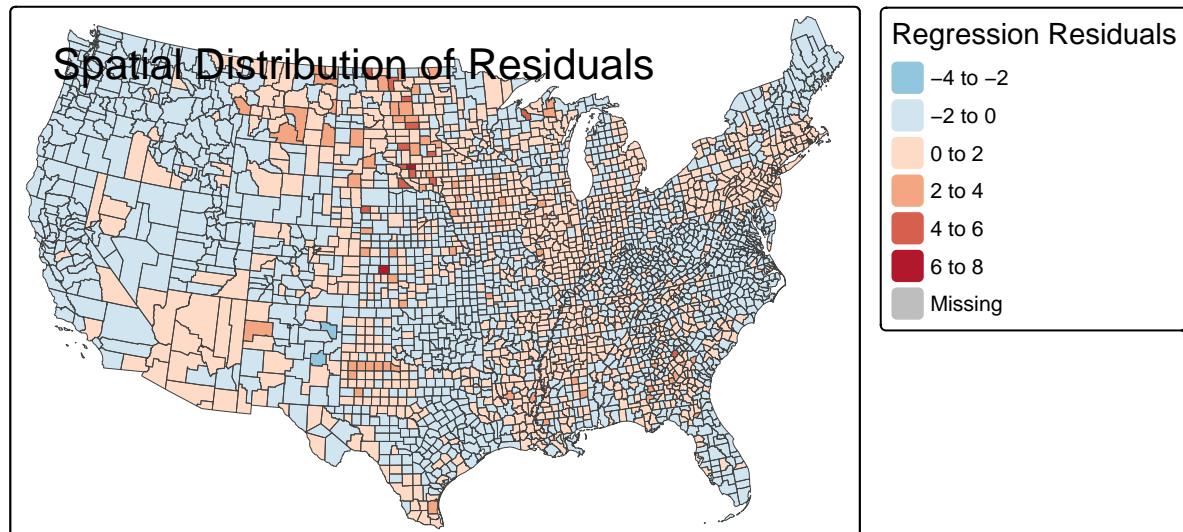
```

```
##
```

```

## -- tmap v3 code detected -----
## [v3->v4] 'tm_tm_fill()': migrate the argument(s) related to the scale of the
## visual variable 'fill' namely 'palette' (rename to 'values') to fill.scale =
## tm_scale(<HERE>).[v3->v4] 'tm_fill()': migrate the argument(s) related to the legend of the
## visual variable 'fill' namely 'title' to 'fill.legend = tm_legend(<HERE>)'[v3->v4] 'tm_layout()': us
## [scale] tm_polygons(): the data variable assigned to 'fill' contains positive and negative values, s
## [cols4all] color palettes: use palettes from the R package cols4all. Run
## 'cols4all::c4a_gui()' to explore them. The old palette name "-RdBu" is named
## "rd_bu" (in long format "brewer.rd_bu")

```



```
moran.test(covid_data$resid, lw, zero.policy = TRUE)
```

```

##
## Moran I test under randomisation
##
## data: covid_data$resid
## weights: lw
## n reduced by no-neighbour observations
##
## Moran I statistic standard deviate = 30.378, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.3264183321    -0.0003220612     0.0001156865

```

To evaluate whether spatial autocorrelation remains after accounting for explanatory variables, we examined the residuals from the multiple linear regression. A spatial map of the residuals reveals clear regional patterns, suggesting potential spatial structure not captured by the model.

This was confirmed by Moran's I test, which indicated significant positive spatial autocorrelation in the residuals ( $I = 0.326$ ,  $p < 0.001$ ). This implies that the model's residuals are not randomly distributed and that spatial dependencies persist. Therefore, a spatial regression model (such as a spatial lag or spatial error model) may be more appropriate for accurately modeling the COVID-19 death rate at the county level.

```
formula <- cases_per_ ~ pct_poc + pct_smoker + pct_povert + pct_obsese + pct_65plus + per_dem

lag_model <- lagsarlm(formula, data = covid_data, listw = lw, zero.policy = TRUE)

error_model <- errorsarlm(formula, data = covid_data, listw = lw, zero.policy = TRUE)

summary(lag_model)
```

**Run spatial lag and spatial error models. Interpret the output from these and show if one of these models fits the data better than the other.**

```
##
## Call:
## lagsarlm(formula = formula, data = covid_data, listw = lw, zero.policy = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.79084 -0.46531 -0.11704  0.35582  6.39417
##
## Type: lag
## Regions with no neighbours included:
## 1771 3033
## Coefficients: (asymptotic standard errors)
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.0054687 0.0145921 0.3748 0.7078
## pct_poc     0.1809474 0.0247723 7.3044 2.784e-13
## pct_smoker  0.0122800 0.0200300 0.6131 0.5398
## pct_povert   0.0883911 0.0176683 5.0028 5.650e-07
## pct_obsese  0.0732371 0.0173724 4.2157 2.490e-05
## pct_65plus   0.0932635 0.0174022 5.3593 8.355e-08
## per_dem     -0.0900365 0.0192713 -4.6720 2.982e-06
##
## Rho: 0.52385, LR test value: 605.5, p-value: < 2.22e-16
## Asymptotic standard error: 0.020435
##      z-value: 25.635, p-value: < 2.22e-16
## Wald statistic: 657.17, p-value: < 2.22e-16
##
## Log likelihood: -3853.77 for lag model
## ML residual variance (sigma squared): 0.66053, (sigma: 0.81273)
## Number of observations: 3108
## Number of parameters estimated: 9
## AIC: 7725.5, (AIC for lm: 8329)
```

```

## LM test for residual autocorrelation
## test value: 23.506, p-value: 1.2452e-06

summary(error_model)

##
## Call:errorsarlm(formula = formula, data = covid_data, listw = lw,
## zero.policy = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.68261 -0.46172 -0.11085  0.33950  6.36561
##
## Type: error
## Regions with no neighbours included:
## 1771 3033
## Coefficients: (asymptotic standard errors)
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.0100684 0.0318838 0.3158 0.752168
## pct_poc     0.3150291 0.0369604 8.5234 < 2.2e-16
## pct_smoker   0.0019546 0.0288556 0.0677 0.945994
## pct_povert    0.1340235 0.0218652 6.1295 8.814e-10
## pct_obsese   0.0866651 0.0276598 3.1333 0.001729
## pct_65plus    0.1326076 0.0212619 6.2369 4.464e-10
## per_dem      -0.1779322 0.0277911 -6.4025 1.529e-10
##
## Lambda: 0.5462, LR test value: 631.3, p-value: < 2.22e-16
## Asymptotic standard error: 0.020724
## z-value: 26.356, p-value: < 2.22e-16
## Wald statistic: 694.65, p-value: < 2.22e-16
##
## Log likelihood: -3840.868 for error model
## ML residual variance (sigma squared): 0.65135, (sigma: 0.80706)
## Number of observations: 3108
## Number of parameters estimated: 9
## AIC: 7699.7, (AIC for lm: 8329)

```

```
AIC(lag_model, error_model)
```

	df	AIC
lag_model	9	7725.541
error_model	9	7699.737

To address residual spatial autocorrelation observed in the OLS model, both spatial lag and spatial error models were estimated. The spatial lag model incorporates spatial dependence in the dependent variable via a spatially lagged term, while the spatial error model accounts for spatial dependence in the error term.

In the spatial lag model, the spatial autoregressive coefficient ( $\rho = 0.524$ ) was highly significant ( $p < 0.001$ ), indicating that COVID-19 death rates in a county are positively influenced by rates in neighboring counties. Key predictors such as the percent of people of color, poverty rate, obesity rate, and elderly population remained significant. The AIC for this model was 7725.5, a notable improvement over the OLS baseline (AIC = 8329).

The spatial error model also showed strong performance. The spatial error coefficient ( $\lambda = 0.546$ ) was significant ( $p < 0.001$ ), suggesting unobserved spatial processes affect the residuals. This model had a slightly lower AIC of 7699.7, indicating a better overall fit than both the lag and OLS models.

Both models outperform the standard OLS regression. However, based on the AIC comparison and slightly lower residual variance, the spatial error model provides the best fit. This suggests that spatial correlation in the data primarily arises from omitted spatially structured variables rather than direct spatial interaction in the dependent variable.

Use the zipped stlouis.shp file and the packages sf, sp, spdep, spgwr & tmap and any others you might need for spatial analysis and mapping.

```
stlouis <- st_read("stlouis/stlouis.shp")
```

Run a non-spatial regression of the homicide rate (HR8893) on police expenditure (PE87) and a local deprivation index (RDAC90) and interpret the output.

```
## Reading layer 'stlouis' from data source
##   '/Users/liuhaojie/Desktop/STATSM222/HW5/stlouis/stlouis.shp'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 78 features and 23 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: -92.70068 ymin: 36.88181 xmax: -87.91657 ymax: 40.32957
## Geodetic CRS:  WGS 84
```

```
model_ols <- lm(HR8893 ~ PE87 + RDAC90, data = stlouis)
```

```
summary(model_ols)
```

```
##
## Call:
## lm(formula = HR8893 ~ PE87 + RDAC90, data = stlouis)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.8749 -2.7648 -0.6719  2.1715 20.2329
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.03748   1.54263   0.024   0.981
## PE87        1.56705   0.37128   4.221 6.75e-05 ***
## RDAC90      5.29091   0.82492   6.414 1.14e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.692 on 75 degrees of freedom
## Multiple R-squared:  0.4065, Adjusted R-squared:  0.3906
## F-statistic: 25.68 on 2 and 75 DF,  p-value: 3.194e-09
```

The results of the non-spatial regression indicate that both police expenditure and local deprivation are significant predictors of neighborhood-level homicide rates in St. Louis. The model explains approximately 39% of the variation in homicide rates, suggesting a moderately strong fit. Specifically, higher levels of police expenditure are associated with higher homicide rates, which may reflect reactive resource allocation to high-crime areas rather than a causal effect. Additionally, local deprivation has a strong and statistically significant positive association with homicide, indicating that more socioeconomically deprived neighborhoods tend to experience substantially higher levels of violence. These findings underscore the importance of underlying structural conditions, such as poverty and inequality, in shaping patterns of violent crime.

```
stlouis_sp <- as(stlouis, "Spatial")
coords <- coordinates(stlouis_sp)
bw <- gwr.sel(HR8893 ~ PE87 + RDAC90, data = stlouis_sp, coords = coords)
```

Use the R package spgwr to show how the R<sup>2</sup> and the partial regression coefficients in your model vary across your study region. Report and interpret all your output.

```
## Warning in gwr.sel(HR8893 ~ PE87 + RDAC90, data = stlouis_sp, coords = coords):
## data is Spatial* object, ignoring coords argument

## Bandwidth: 193.9725 CV score: 2541.237
## Bandwidth: 313.5408 CV score: 2535.366
## Bandwidth: 387.438 CV score: 2531.794
## Bandwidth: 433.1091 CV score: 2530.234
## Bandwidth: 461.3353 CV score: 2529.453
## Bandwidth: 478.7801 CV score: 2529.028
## Bandwidth: 489.5615 CV score: 2528.784
## Bandwidth: 496.2248 CV score: 2528.64
## Bandwidth: 500.343 CV score: 2528.554
## Bandwidth: 502.8881 CV score: 2528.502
## Bandwidth: 504.4611 CV score: 2528.47
## Bandwidth: 505.4333 CV score: 2528.45
## Bandwidth: 506.0341 CV score: 2528.438
## Bandwidth: 506.4055 CV score: 2528.43
## Bandwidth: 506.635 CV score: 2528.426
## Bandwidth: 506.7768 CV score: 2528.423
## Bandwidth: 506.8645 CV score: 2528.421
## Bandwidth: 506.9186 CV score: 2528.42
## Bandwidth: 506.9521 CV score: 2528.419
## Bandwidth: 506.9728 CV score: 2528.419
## Bandwidth: 506.9856 CV score: 2528.419
## Bandwidth: 506.9935 CV score: 2528.419
## Bandwidth: 506.9984 CV score: 2528.419
## Bandwidth: 507.0014 CV score: 2528.418
## Bandwidth: 507.0033 CV score: 2528.418
## Bandwidth: 507.0044 CV score: 2528.418
## Bandwidth: 507.0051 CV score: 2528.418
## Bandwidth: 507.0056 CV score: 2528.418
## Bandwidth: 507.0059 CV score: 2528.418
## Bandwidth: 507.006 CV score: 2528.418
```

```

## Bandwidth: 507.0061 CV score: 2528.418
## Bandwidth: 507.0062 CV score: 2528.418
## Bandwidth: 507.0062 CV score: 2528.418
## Bandwidth: 507.0062 CV score: 2528.418

## Warning in gwr.sel(HR8893 ~ PE87 + RDAC90, data = stlouis_sp, coords = coords):
## Bandwidth converged to upper bound: 507.006301412274

# Fit GWR model
gwr_model <- gwr(HR8893 ~ PE87 + RDAC90,
                  data = stlouis_sp,
                  coords = coords,
                  bandwidth = bw,
                  hatmatrix = TRUE,
                  se.fit = TRUE)

## Warning in gwr(HR8893 ~ PE87 + RDAC90, data = stlouis_sp, coords = coords, :
## data is Spatial* object, ignoring coords argument

gwr_model

## Call:
## gwr(formula = HR8893 ~ PE87 + RDAC90, data = stlouis_sp, coords = coords,
##      bandwidth = bw, hatmatrix = TRUE, se.fit = TRUE)
## Kernel function: gwr.Gauss
## Fixed bandwidth: 507.0062
## Summary of GWR coefficient estimates at data points:
##          Min.    1st Qu.   Median    3rd Qu.   Max. Global
## X.Intercept. -0.100194 -0.050197 -0.018489  0.018239  0.074903 0.0375
## PE87         1.575635  1.587146  1.594174  1.601212  1.610909 1.5670
## RDAC90       5.244074  5.311522  5.354990  5.395810  5.453493 5.2909
## Number of data points: 78
## Effective number of parameters (residual: 2traceS - traceS'S): 3.316572
## Effective degrees of freedom (residual: 2traceS - traceS'S): 74.68343
## Sigma (residual: 2traceS - traceS'S): 4.691573
## Effective number of parameters (model: traceS): 3.161564
## Effective degrees of freedom (model: traceS): 74.83844
## Sigma (model: traceS): 4.686712
## Sigma (ML): 4.590747
## AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 468.018
## AIC (GWR p. 96, eq. 4.22): 462.2666
## Residual sum of squares: 1643.847
## Quasi-global R2: 0.4091174

gwr_output <- as.data.frame(gwr_model$SDF)

stlouis$gwr_r2      <- gwr_output$localR2
stlouis$coef_PE87    <- gwr_output$PE87
stlouis$coef_RDAC90  <- gwr_output$RDAC90
stlouis

## Simple feature collection with 78 features and 26 fields

```

```

## Geometry type: POLYGON
## Dimension: XY
## Bounding box: xmin: -92.70068 ymin: 36.88181 xmax: -87.91657 ymax: 40.32957
## Geodetic CRS: WGS 84
## First 10 features:
##   NAME STATE_NAME STATE_FIPS CNTY_FIPS FIPS STFIPS COFIPS FIPSNO HR7984
## 1 Logan Illinois 17 107 17107 17 107 17107 2.115428
## 2 Adams Illinois 17 001 17001 17 1 17001 4.464496
## 3 Menard Illinois 17 129 17129 17 129 17129 4.307312
## 4 Cass Illinois 17 017 17017 17 17 17017 2.258866
## 5 Brown Illinois 17 009 17009 17 9 17009 5.935246
## 6 Macon Illinois 17 115 17115 17 115 17115 3.613635
## 7 Sangamon Illinois 17 167 17167 17 167 17167 5.774120
## 8 Marion Missouri 29 127 29127 29 127 29127 1.742342
## 9 Morgan Illinois 17 137 17137 17 137 17137 3.124540
## 10 Pike Illinois 17 149 17149 17 149 17149 2.643242
##   HR8488 HR8893 HC7984 HC8488 HC8893 P07984 P08488 P08893 PE77
## 1 1.290722 1.624458 4 2 3 189087 154952 184677 5.104320
## 2 2.655839 2.255492 19 9 9 425580 338876 399026 2.304996
## 3 1.742433 1.467890 3 1 1 69649 57391 68125 5.183402
## 4 1.437029 2.484256 2 1 2 88540 69588 80507 3.955886
## 5 0.000000 0.000000 2 0 0 33697 28462 35051 5.755396
## 6 6.036815 9.048673 28 37 64 774843 612906 707286 4.818652
## 7 5.441418 6.029489 61 48 65 1056438 882123 1078035 5.444019
## 8 0.000000 1.800385 3 0 3 172182 140910 166631 6.758459
## 9 2.166249 4.581251 7 4 10 224033 184651 218281 6.651120
## 10 3.298008 3.790607 3 3 4 113497 90964 105524 2.861865
##   PE82 PE87 RDAC80 RDAC85 RDAC90
## 1 6.595780 5.832951 -0.991256 -0.940265 -0.845005
## 2 4.255254 5.457145 -0.509511 -0.391588 -0.271549
## 3 3.012367 2.316253 -1.340772 -1.114002 -0.859035
## 4 2.263292 5.263547 -0.754894 -0.511259 -0.276122
## 5 4.728186 4.653891 0.096642 -0.130929 -0.305658
## 6 4.947508 5.131238 -0.664045 -0.437859 -0.229741
## 7 5.802772 6.218296 -0.790212 -0.689591 -0.586439
## 8 4.272589 4.413135 -0.409684 -0.210570 -0.016762
## 9 6.023764 7.330383 -0.477808 -0.621812 -0.733578
## 10 3.352330 3.189221 0.094225 -0.064360 -0.185323
##   geometry gwr_r2 coef_PE87 coef_RDAC90
## 1 POLYGON ((-89.58522 39.9787..., 0.4182092 1.575635 5.453493
## 2 POLYGON ((-90.92154 39.8474..., 0.4178916 1.579577 5.409474
## 3 POLYGON ((-89.9976 39.9062,... 0.4181063 1.577661 5.439998
## 4 POLYGON ((-90.58536 39.8807..., 0.4180207 1.578994 5.428259
## 5 POLYGON ((-90.57704 39.8454..., 0.4179442 1.579669 5.417249
## 6 POLYGON ((-89.0322 39.65652..., 0.4178031 1.578249 5.446742
## 7 POLYGON ((-89.70396 39.5280..., 0.4177294 1.580950 5.428045
## 8 POLYGON ((-91.85072 39.6611..., 0.4175124 1.582083 5.390261
## 9 POLYGON ((-90.1548 39.52558..., 0.4176632 1.582273 5.415184
## 10 POLYGON ((-91.20325 39.6000..., 0.4174332 1.584168 5.395907

```

The Geographically Weighted Regression (GWR) analysis reveals that the effects of police expenditure (PE87) and local deprivation (RDAC90) on homicide rates (HR8893) vary modestly across the study region. Using a fixed bandwidth of approximately 507—determined via cross-validation—the model produced a quasi-global \$R^2\$ of 0.409, closely matching the global OLS model but allowing for spatial variation in

coefficient estimates. Across neighborhoods, local  $R^2$  values remained stable (around 0.418), indicating a fairly consistent model fit throughout the region.

The spatial distribution of the coefficients shows that both predictors have a uniformly positive relationship with homicide, but with small variations in magnitude. The effect of police expenditure ranged from about 1.58 to 1.61, suggesting that in all areas, higher spending is associated with higher homicide rates—likely reflecting reactive spending patterns rather than prevention. The coefficient for local deprivation ranged from about 5.24 to 5.45, consistently indicating that more deprived neighborhoods experience higher levels of violent crime. These results underscore that while the direction of influence is stable, GWR helps uncover subtle spatial nuances that a global model would miss.