

Tarea 4: Crimen en Cape Town

Amparo Galvez Vilar

2024-12-07

Contents

Carga de librerías	1
Lectura de datos	1
Estructura y características generales de los datos	1
Análisis descriptivo	6
Visualización	8
Modelo estadístico	11
Incorporación de estructura espacial	16
Evaluación de la Dependencia Espacial: Índice de Moran's I	16
Ajuste de un Modelo Autorregresivo Espacial: Modelo CAR	17
Comparación: Diferencias entre Modelos	19
Análisis Específico de Alta Criminalidad	19
Clasificación de Barrios	19
Modelo Logístico	20
Predicción de Alta Criminalidad	22
Mapa de Predicciones	23
Conclusión	24

Carga de librerías

```
library(sf)
library(dplyr)
library(spdep)
library(ggplot2)
library(spatialreg)
```

Lectura de datos

```
datos = st_read('./CapeTown.gpkg', layer='LatticeDataR', quiet=T)
```

Estructura y características generales de los datos

Dado que no disponemos del número de personas que viven en cada barrio y no tenemos acceso a la oficina de estadística, podemos asumir que el número de habitantes en cada barrio es el total de habitantes de Ciudad

del Cabo dividido entre 777, que es el número de barrios. Esta suposición se basa en la falta de información específica sobre la población de cada barrio.

```
# Explorar las variables
# Ver la estructura de los datos
str(datos)
```

```
## Classes 'sf' and 'data.frame': 777 obs. of 26 variables:
## $ OBJECTID          : num 776 560 555 472 223 215 522 367 329 554 ...
## $ OFC_SBRB_N        : chr "AAN DE WIJNLANDEN ESTATE" "ACACIA PARK" "ADMIRALS PARK" "ADRIAANSE" ...
## $ SHAPE_Leng         : num 4534 8813 2208 2888 1585 ...
## $ SHAPE_Area         : num 513799 2834563 214057 432164 92891 ...
## $ OFFICIAL_SUBURB   : chr NA "ACACIA PARK" "ADMIRALS PARK" "ADRIAANSE" ...
## $ NUM_RES_PROP       : num NA NA 401 378 NA NA 102 351 291 400 ...
## $ MED_LAND_EXTENT.m2 : num NA NA 235 127 NA ...
## $ MED_TOT_BLD_AREA.m2: num NA NA 111 65 NA NA 310 264 224 158 ...
## $ GV2015VAL          : num NA NA 760000 470000 NA NA 2000000 1450000 1200000 1000000 ...
## $ X                  : num 291192 270800 304336 277012 280480 ...
## $ Y                  : num 6234420 6246596 6219347 6241830 6235466 ...
## $ AVESTADist         : num 8330 3345 1404 2172 2913 ...
## $ AVAPREArea         : num 2.02e+08 1.46e+07 5.62e+07 1.38e+07 7.38e+06 ...
## $ AvePrecinctsLen    : num 84528 23641 68684 20945 15708 ...
## $ ROBRes             : num 33.5 26.5 9 15.3 134.5 ...
## $ BUGRes              : num 882 329 234 323 454 ...
## $ CRIME               : num 916 356 243 338 588 ...
## $ AVMEDINC            : num 57601 86401 115201 14400 14400 ...
## $ BACH                : chr "690" "2249, 1430" "61" "54" ...
## $ AVBACH              : num 690 1840 61 54 44 ...
## $ STA                 : chr "KLEINVLEI, STELLENBOSCH" "GOODWOOD, KENSINGTON, MAITLAND, MILNERTON" ...
## $ STAD                : chr "3055.2262446575687, 13604.800887177007" "4010.592536839889, 1789.231299 ...
## $ PREA                : chr "26648462.479252875, 376867651.08642346" "15287334.022337114, 6062179.3 ...
## $ PREL                : chr "24674.257370602747, 144381.1457476814" "18820.679406607203, 13920.7614 ...
## $ WARD                : chr "14" "55, 56" "100" "42" ...
## $ geom                :sfc_MULTIPOINT of length 777; first list element: List of 1
##   ..$ :List of 1
##     ...$ : num [1:32, 1:2] 291525 291415 291237 291236 291046 ...
##     ..- attr(*, "class")= chr [1:3] "XY" "MULTIPOLYGON" "sfg"
##     - attr(*, "sf_column")= chr "geom"
##     - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",...: NA NA NA NA NA NA NA NA NA ...
##     ..- attr(*, "names")= chr [1:25] "OBJECTID" "OFC_SBRB_N" "SHAPE_Leng" "SHAPE_Area" ...
# Resumen general de los datos
summary(datos)
```

```
##      OBJECTID    OFC_SBRB_N        SHAPE_Leng      SHAPE_Area
## Min.   : 1   Length:777      Min.   : 590.9   Min.   : 17814
## 1st Qu.:195  Class :character  1st Qu.: 2490.2   1st Qu.: 285620
## Median :389   Mode  :character  Median : 3923.7   Median : 675216
## Mean   :389                    Mean   : 6148.8   Mean   : 3180585
## 3rd Qu.:583                    3rd Qu.: 5969.3   3rd Qu.: 1448977
## Max.   :777                    Max.   :179726.3  Max.   :403341748
##
##      OFFICIAL_SUBURB    NUM_RES_PROP    MED_LAND_EXTENT.m2 MED_TOT_BLD_AREA.m2
## Length:777      Min.   : 1.0   Min.   : 36.0   Min.   : 20.0
## Class :character  1st Qu.: 152.0  1st Qu.: 285.5  1st Qu.: 102.5
## Mode  :character  Median : 407.0  Median : 496.0  Median : 188.0
```

```

##          Mean   : 931.2    Mean   : 3186.5    Mean   : 462.0
##          3rd Qu.: 1022.0   3rd Qu.: 991.5    3rd Qu.: 288.5
##          Max.   :18683.0   Max.   :944565.0   Max.   :6655.0
##          NA's   :142      NA's   :142      NA's   :142
##  GV2015VAL           X             Y             AVESTADist
##  Min.   : 56000    Min.   :254493    Min.   :6205787   Min.   : 66.14
##  1st Qu.: 490000   1st Qu.:274964    1st Qu.:6230033   1st Qu.: 1525.77
##  Median : 1050000  Median :282974    Median :6241387   Median : 2299.39
##  Mean   : 1497628   Mean   :282980    Mean   :6240142   Mean   : 2719.02
##  3rd Qu.: 1900000  3rd Qu.:288651    3rd Qu.:6247978   3rd Qu.: 3253.86
##  Max.   :19050000  Max.   :308766    Max.   :6288870   Max.   :24072.59
##  NA's   :142
##  AVAPREArea        AvePrecinctsLen    ROBRes       BUGRes
##  Min.   : 5938870  Min.   : 9349    Min.   : 0.00  Min.   : 0.0
##  1st Qu.: 18503824 1st Qu.: 22473   1st Qu.: 17.50 1st Qu.: 401.0
##  Median : 31312269  Median : 32287   Median : 28.00  Median : 700.0
##  Mean   : 70750548  Mean   : 39914   Mean   : 31.17  Mean   : 673.5
##  3rd Qu.:1075556893 3rd Qu.: 62382   3rd Qu.: 35.50 3rd Qu.: 978.5
##  Max.   :507786246  Max.   :128914   Max.   :156.50  Max.   :1177.0
##
##  CRIME            AVMEDINC        BACH         AVBACH
##  Min.   : 0.0     Min.   :14400    Length:777    Min.   : 44.0
##  1st Qu.: 428.0   1st Qu.: 57601   Class :character 1st Qu.: 270.0
##  Median : 733.5   Median :115201   Mode  :character  Median : 716.5
##  Mean   : 704.6   Mean   :127986                    Mean   : 781.8
##  3rd Qu.:1012.0   3rd Qu.:187201                    3rd Qu.:1233.0
##  Max.   :1220.0   Max.   :460801                    Max.   :2511.0
##
##  STA              STAD           PREA         PREL
##  Length:777       Length:777      Length:777    Length:777
##  Class :character Class :character  Class :character Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##  WARD             geom
##  Length:777       MULTIPOLYGON :777
##  Class :character epsg:32734   : 0
##  Mode  :character +proj=utm ...: 0
##
##
```

El conjunto de datos contiene un total de 777 observaciones y 26 variables, configurándose como un conjunto de datos espaciales (sf). Estas variables incluyen tipos diversos como numéricas, categóricas y geométricas, donde la columna geom contiene polígonos.

La variable que estamos observando es de tipo flotante (CRIME), lo cual es inesperado, ya que lo esperado sería un número entero. Este valor flotante podría corresponder a una media de varios años. Para simplificar, redondeamos este valor a un entero. Así, podemos continuar con un análisis más manejable y consistente.

```
# Redondear la variable CRIME a enteros
datos$CRIME <- round(datos$CRIME)
```

Si una variable tiene muchos NAs, lo mejor es descartarla porque su información es muy deficiente.

Variables con muchos valores faltantes:

- OFFICIAL_SUBURB: 2.96% de valores faltantes.
- NUM_RES_PROP: 18.28% de valores faltantes.
- MED_LAND_EXTENT.m2: 18.28% de valores faltantes.
- MED_TOT_BLD_AREA.m2: 18.28% de valores faltantes.
- GV2015VAL: 18.28% de valores faltantes.

Dado que no se cuenta con la posibilidad de imputar valores faltantes, es recomendable eliminar la variable GV2015VAL del análisis para asegurar la calidad del estudio. Las variables con un alto porcentaje de valores faltantes, como GV2015VAL, pueden introducir sesgos y afectar la integridad de los resultados. Además, mantener variables con información incompleta podría conducir a conclusiones erróneas y reducir la confiabilidad del análisis. Por lo tanto, eliminarlas permitirá centrarse en las variables con datos completos y precisos, mejorando la robustez y la precisión del estudio.

```
# Descartar variables con muchos valores faltantes
datos <- datos %>% select(-OFFICIAL_SUBURB, -NUM_RES_PROP, -MED_LAND_EXTENT.m2, -MED_TOT_BLD_AREA.m2, -GV2015VAL)

str(datos)

## Classes 'sf' and 'data.frame': 777 obs. of 21 variables:
## $ OBJECTID      : num  776 560 555 472 223 215 522 367 329 554 ...
## $ OFC_SBRB_N    : chr "AAN DE WIJNLANDEN ESTATE" "ACACIA PARK" "ADMIRALS PARK" "ADRIAANSE" ...
## $ SHAPE_Leng     : num 4534 8813 2208 2888 1585 ...
## $ SHAPE_Area    : num 513799 2834563 214057 432164 92891 ...
## $ X              : num 291192 270800 304336 277012 280480 ...
## $ Y              : num 6234420 6246596 6219347 6241830 6235466 ...
## $ AVESTADist    : num 8330 3345 1404 2172 2913 ...
## $ AVAPREArea    : num 2.02e+08 1.46e+07 5.62e+07 1.38e+07 7.38e+06 ...
## $ AvePrecinctsLen: num 84528 23641 68684 20945 15708 ...
## $ ROBRes        : num 33.5 26.5 9 15.3 134.5 ...
## $ BUGRes        : num 882 329 234 323 454 ...
## $ CRIME          : num 916 356 243 338 588 ...
## $ AVMEDINC      : num 57601 86401 115201 14400 14400 ...
## $ BACH           : chr "690" "2249, 1430" "61" "54" ...
## $ AVBACH         : num 690 1840 61 54 44 ...
## $ STA            : chr "KLEINVLEI, STELLENBOSCH" "GOODWOOD, KENSINGTON, MAITLAND, MILNERTON" "GORDONSTON, DUNDEE" ...
## $ STAD           : chr "3055.2262446575687, 13604.800887177007" "4010.592536839889, 1789.2312958013" ...
## $ PREA           : chr "26648462.479252875, 376867651.08642346" "15287334.022337114, 6062179.348738" ...
## $ PREL           : chr "24674.257370602747, 144381.1457476814" "18820.679406607203, 13920.761472330" ...
## $ WARD           : chr "14" "55, 56" "100" "42" ...
## $ geom           :sfc_MULTIPOINT of length 777; first list element: List of 1
##   ..$ :List of 1
##     ...$ : num [1:32, 1:2] 291525 291415 291237 291236 291046 ...
##     ...- attr(*, "class")= chr [1:3] "XY" "MULTIPOINT" "sfg"
##     - attr(*, "sf_column")= chr "geom"
##     - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",...: NA NA NA NA NA NA NA NA NA ...
##     ..- attr(*, "names")= chr [1:20] "OBJECTID" "OFC_SBRB_N" "SHAPE_Leng" "SHAPE_Area" ...
summary(datos)

##      OBJECTID      OFC_SBRB_N      SHAPE_Leng      SHAPE_Area
## Min.    : 1      Length:777      Min.    : 590.9      Min.    : 17814
```

```

## 1st Qu.:195 Class :character 1st Qu.: 2490.2 1st Qu.: 285620
## Median :389 Mode :character Median : 3923.7 Median : 675216
## Mean :389 Mean : 6148.8 Mean : 3180585
## 3rd Qu.:583 3rd Qu.: 5969.3 3rd Qu.: 1448977
## Max. :777 Max. :179726.3 Max. :403341748
##          X           Y           AVESTADist      AVAPREArea
## Min. :254493  Min. :6205787  Min. : 66.14  Min. : 5938870
## 1st Qu.:274964 1st Qu.:6230033 1st Qu.: 1525.77 1st Qu.: 18503824
## Median :282974  Median :6241387  Median : 2299.39  Median : 31312269
## Mean :282980   Mean :6240142  Mean : 2719.02  Mean : 70750548
## 3rd Qu.:288651 3rd Qu.:6247978 3rd Qu.: 3253.86 3rd Qu.:107556893
## Max. :308766   Max. :6288870  Max. :24072.59  Max. :507786246
## AvePrecinctsLen      ROBRes      BUGRes      CRIME
## Min. : 9349  Min. : 0.00  Min. : 0.0  Min. : 0.0
## 1st Qu.: 22473 1st Qu.: 17.50 1st Qu.: 401.0 1st Qu.: 428.0
## Median : 32287 Median : 28.00 Median : 700.0 Median : 734.0
## Mean : 39914  Mean : 31.17 Mean : 673.5 Mean : 704.7
## 3rd Qu.: 62382 3rd Qu.: 35.50 3rd Qu.: 978.5 3rd Qu.:1012.0
## Max. :128914  Max. :156.50 Max. :1177.0 Max. :1220.0
##          AVMEDINC      BACH      AVBACH      STA
## Min. : 14400 Length:777  Min. : 44.0 Length:777
## 1st Qu.: 57601 Class :character 1st Qu.: 270.0 Class :character
## Median :115201 Mode :character Median : 716.5 Mode :character
## Mean :127986  Mean : 781.8
## 3rd Qu.:187201 3rd Qu.:1233.0
## Max. :460801  Max. :2511.0
##          STAD      PREA      PREL      WARD
## Length:777 Length:777 Length:777 Length:777
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
##          geom
## MULTIPOLYGON :777
## epsg:32734 : 0
## +proj=utm ...: 0
##
##
##

```

Se han mantenido las variables clave para el análisis, especialmente aquellas relacionadas con la variable de interés, CRIME. La eliminación de las variables con muchos valores faltantes ha mejorado la calidad general del conjunto de datos. Esto asegura que el análisis se base en datos completos y más confiables, lo cual es crucial para evitar sesgos y errores en los resultados.

Por otro lado, las variables numéricas restantes, como SHAPE_Leng, SHAPE_Area y AVMEDINC, presentan una buena dispersión y un rango de valores amplio. Esta característica es útil para un análisis detallado. Asimismo, variables como OFC_SBRB_N, STA, STAD, PREA, PREL y WARD proporcionan una clasificación que puede ser útil en análisis más detallados o segmentaciones.

Al mantener variables críticas y eliminar aquellas con muchos valores faltantes, el análisis se vuelve más robusto y preciso. Esto permitirá identificar patrones y relaciones significativas. En conclusión, este proceso de limpieza y reducción de datos es esencial para obtener resultados válidos y confiables en el análisis, especialmente cuando se trata de variables complejas y relaciones múltiples.

Análisis descriptivo

Los barrios se clasificaron en dos categorías: alta criminalidad y baja criminalidad. Los barrios con más de 500 crímenes se consideraron de alta criminalidad, mientras que aquellos con 500 o menos crímenes se consideraron de baja criminalidad. Para realizar esta clasificación, se añadió una nueva columna binaria llamada “Criminality”, la cual especifica si un barrio pertenece a la categoría de alta o baja criminalidad.

Esta columna permite realizar análisis diferenciados por nivel de criminalidad y explorar cómo se comportan las variables en cada grupo.

```
# Crear la columna de clasificación de criminalidad
datos <- datos %>%
  mutate(Criminality = ifelse(CRIME > 500, "Alta", "Baja"))
head(datos)

## Simple feature collection with 6 features and 21 fields
## Geometry type: MULTIPOLYGON
## Dimension: XY
## Bounding box: xmin: 268986.2 ymin: 6218935 xmax: 304690.9 ymax: 6247833
## Projected CRS: WGS 84 / UTM zone 34S
##   OBJECTID          OFC_SBRB_N SHAPE_Leng SHAPE_Area      X      Y
## 1    776 AAN DE WIJNLANDEN ESTATE  4534.160 513798.70 291191.9 6234420
## 2    560           ACACIA PARK  8813.434 2834563.17 270800.0 6246596
## 3    555        ADMIRALS PARK  2208.167 214056.57 304335.7 6219347
## 4    472        ADRIAANSE     2887.830 432164.26 277012.0 6241830
## 5    223         AIRPORT     1585.438 92890.75 280480.2 6235466
## 6    215        AIRPORT CITY  4699.484 865711.27 277716.0 6237285
##   AVESTADist AVAPREArea AvePrecinctsLen    ROBRes    BUGRes CRIME AVMEDINC
## 1  8330.014  201758057       84527.70  33.50000 882.0000  916  57600.5
## 2  3344.736  14589290       23640.66  26.50000 329.0000  356  86400.5
## 3  1403.838  56236915       68684.24   9.00000 234.0000  243 115200.5
## 4  2172.129  13804307       20944.87  15.33333 322.6667  338 14400.5
## 5  2913.229  7377339       15707.71 134.50000 453.5000  588 14400.5
## 6  2821.543  11190647       22269.43 115.50000 598.0000  714 36000.5
##   BACH AVBACH                      STA
## 1    690  690.0          KLEINVLEI, STELENBOSCH
## 2  2249, 1430 1839.5 GOODWOOD, KENSINGTON, MAITLAND, MILNERTON
## 3     61   61.0          GORDONS BAY
## 4     54   54.0        BISHOP LAVIS, ELSIES RIVER, RAVENSMED
## 5     44   44.0          NYANGA, PHILIPPI EAST
## 6   565, 67  316.0        BISHOP LAVIS, NYANGA
##                                     STAD
## 1                               3055.2262446575687, 13604.800887177007
## 2  4010.592536839889, 1789.2312958018001, 4667.737433739719, 2911.3819883366514
## 3                               1403.8380387141265
## 4                               1540.550397846405, 2760.7490902858694, 2215.088920795488
## 5                               3824.896111675613, 2001.561230536149
## 6                               4379.352663422444, 1263.733329244032
##                                     PREA
## 1                               26648462.479252875, 376867651.08642346
## 2  15287334.022337114, 6062179.348738552, 8429338.301907197, 28578308.840154026
## 3                               56236914.62510753
## 4  13565485.01144061, 13122596.287437074, 14724839.638004517
## 5                               8815808.262320677, 5938870.425663566
## 6  13565485.01144061, 8815808.262320677
```

```

##                                     PREL
## 1                               24674.257370602747, 144381.1457476814
## 2 18820.679406607203, 13920.761472336622, 25721.377934366767, 36099.82024568648
## 3                                         68684.23613487242
## 4                               22472.394907592654, 14937.412148219979, 25424.79447680546
## 5                                         22066.464361358834, 9348.958808021645
## 6                               22472.394907592654, 22066.464361358834
##   WARD          geom Criminality
## 1    14 MULTIPOLYGON (((291525.1 62...
## 2    55, 56 MULTIPOLYGON (((270897.9 62...
## 3   100 MULTIPOLYGON (((304475.1 62...
## 4    42 MULTIPOLYGON (((277553.3 62...
## 5    36 MULTIPOLYGON (((280852.4 62...
## 6   22, 40 MULTIPOLYGON (((278100.1 62...

```

A continuación, se calcularon estadísticas descriptivas (media y mediana) para las variables seleccionadas, agrupando los barrios según su nivel de criminalidad. Estas estadísticas ofrecen una visión general de cómo las características de los barrios varían entre los dos grupos:

- CRIME: Número de crímenes.
- AVMEDINC: Renta familiar promedio.
- AVAPREArea: Superficie promedio de los barrios.
- AVBACH: Proporción de habitantes con título universitario.

El código utilizado para realizar este análisis es el siguiente:

```

# Calcular estadísticas descriptivas por grupo de criminalidad
stats <- datos %>%
  group_by(Criminality) %>%
  summarise(
    # Estadísticas de CRIME
    mean_CRIME = mean(CRIME, na.rm = TRUE),
    median_CRIME = median(CRIME, na.rm = TRUE),
    # Estadísticas de AVMEDINC (Renta familiar promedio)
    mean_AVMEDINC = mean(AVMEDINC, na.rm = TRUE),
    median_AVMEDINC = median(AVMEDINC, na.rm = TRUE),
    # Estadísticas de AVAPREArea (Superficie promedio de los barrios)
    mean_AVAPREArea = mean(AVAPREArea, na.rm = TRUE),
    median_AVAPREArea = median(AVAPREArea, na.rm = TRUE),
    # Estadísticas de AVBACH (Proporción de titulados universitarios)
    mean_AVBACH = mean(AVBACH, na.rm = TRUE),
    median_AVBACH = median(AVBACH, na.rm = TRUE)
  )
# Mostrar las estadísticas descriptivas
print(stats)

```

```

## Simple feature collection with 2 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: 251427.6 ymin: 6195136 xmax: 315883.1 ymax: 6293569
## Projected CRS:  WGS 84 / UTM zone 34S
## # A tibble: 2 x 10
##   Criminality mean_CRIME median_CRIME mean_AVMEDINC median_AVMEDINC
##   <chr>           <dbl>        <dbl>        <dbl>        <dbl>
## 1 Alta            858.         866       139511.      115200.

```

```

## 2 Baja          349.          361          101242.          86400.
## # i 5 more variables: mean_AVAPREArea <dbl>, median_AVAPREArea <dbl>,
## #   mean_AVBACH <dbl>, median_AVBACH <dbl>, geom <MULTIPOLYGON [m]>

```

El análisis descriptivo reveló diferencias significativas entre los barrios con alta y baja criminalidad:

1. CRIME (Número de crímenes):

- Los barrios con alta criminalidad tienen un promedio de 857.84 crímenes, comparado con los 349.18 en los barrios con baja criminalidad.
- La diferencia en las medianas (866 frente a 361) refuerza esta segmentación.

2. AVMEDINC (Renta familiar promedio):

- Los barrios de alta criminalidad tienen una renta promedio más elevada (139,510.5) que los de baja criminalidad (101,241.7).
- Este patrón puede sugerir una dinámica socioeconómica inversa a la esperada, que requeriría un análisis adicional.

3. AVAPREArea (Superficie promedio de los barrios):

- La superficie promedio es mayor en barrios de alta criminalidad (86,314,785) en comparación con los de baja criminalidad (34,633,538).
- Esto puede estar relacionado con factores como densidad poblacional o planificación urbana.

4. AVBACH (Proporción de habitantes con título universitario):

- La proporción de titulados universitarios es mayor en barrios de alta criminalidad (842.35) que en los de baja criminalidad (641.36).
- Este hallazgo es interesante y plantea preguntas sobre las relaciones entre educación y criminalidad.

Conclusiones

El análisis descriptivo destaca diferencias importantes entre los barrios según su nivel de criminalidad. Las variables asociadas a renta, superficie y nivel educativo muestran patrones que merecen un análisis más profundo. En particular, la relación positiva entre renta y criminalidad podría ser indicativa de factores contextuales únicos en la región estudiada.

Visualización

En esta sección se exploran visualmente las diferencias entre los barrios con alta y baja criminalidad en relación con las variables clave del análisis.

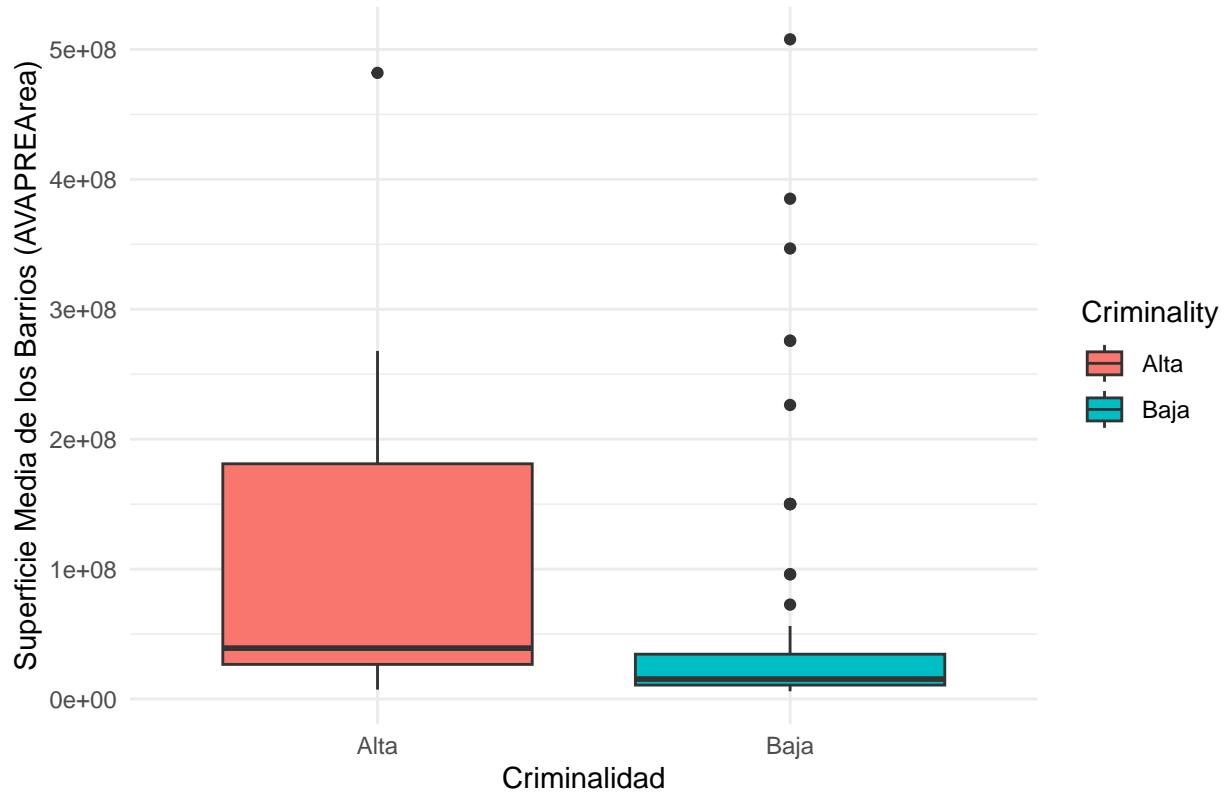
Superficie Media de los Barrios (AVAPREArea)

```

ggplot(datos, aes(x = Criminality, y = AVAPREArea, fill = Criminality)) +
  geom_boxplot() +
  theme_minimal() +
  labs(
    title = "Superficie Media según el Nivel de Criminalidad",
    x = "Criminalidad",
    y = "Superficie Media de los Barrios (AVAPREArea)"
  )

```

Superficie Media según el Nivel de Criminalidad

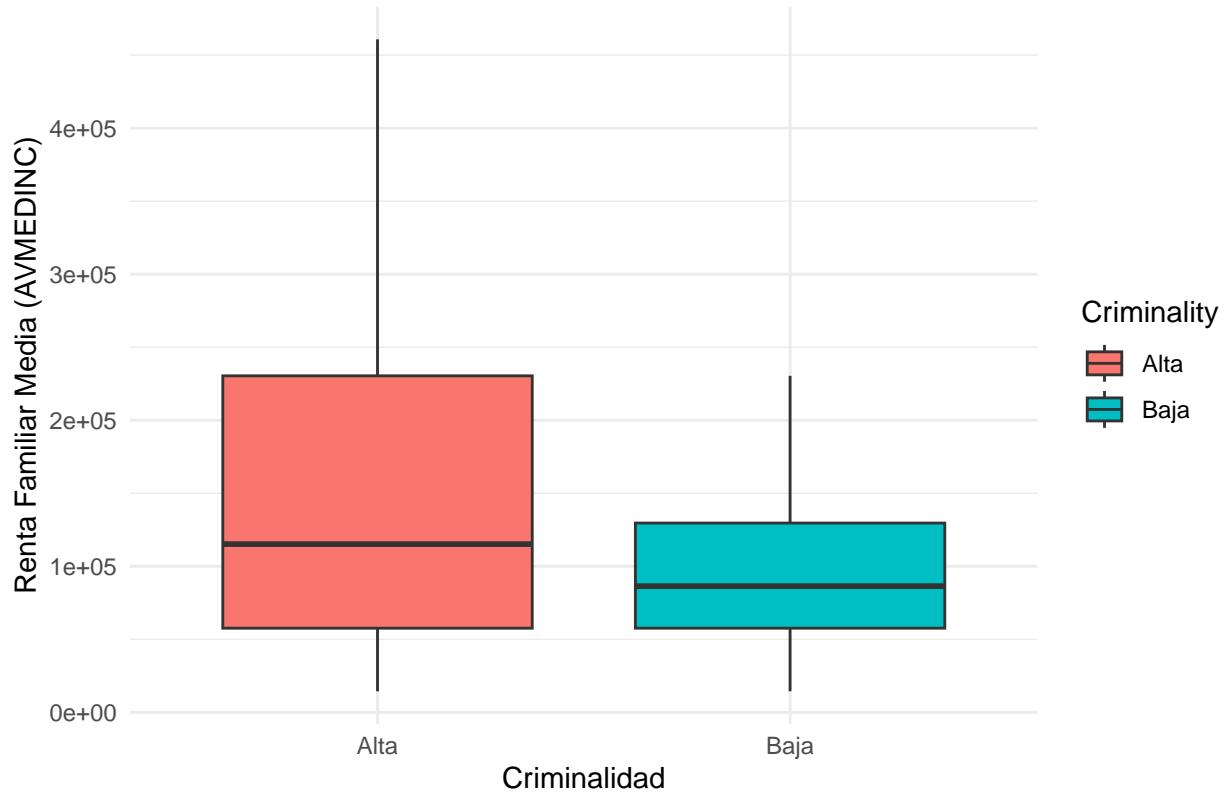


Los barrios con alta criminalidad tienen una mediana de superficie mayor en comparación con los de baja criminalidad. Además, la mayor dispersión observada en estos barrios sugiere que incluyen tanto zonas extensas como áreas más pequeñas. Este resultado podría indicar que los barrios más grandes son más difíciles de vigilar, lo que facilita la comisión de delitos, especialmente en áreas menos vigiladas o periféricas.

Renta Familiar Media (AVMEDINC)

```
ggplot(datos, aes(x = Criminality, y = AVMEDINC, fill = Criminality)) +  
  geom_boxplot() +  
  theme_minimal() +  
  labs(  
    title = "Renta Familiar Media según el Nivel de Criminalidad",  
    x = "Criminalidad",  
    y = "Renta Familiar Media (AVMEDINC)"  
)
```

Renta Familiar Media según el Nivel de Criminalidad

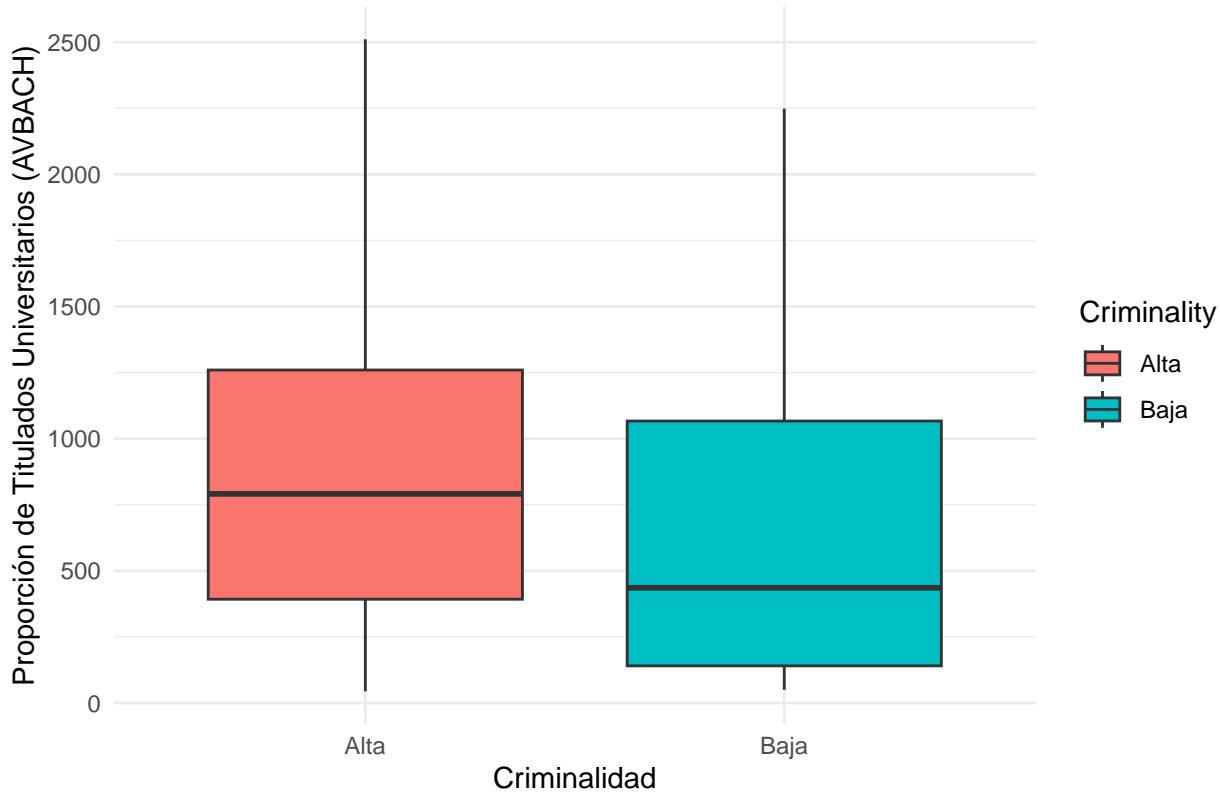


La media y la mediana de la renta familiar son más altas en los barrios con alta criminalidad. Esto sugiere que los delitos económicos, como los robos a casas de alto valor, podrían estar relacionados con estos barrios. Por otro lado, en los barrios con baja criminalidad, la renta promedio es significativamente más baja, lo que podría estar asociado a factores como un menor nivel de urbanización o un acceso limitado a oportunidades económicas.

Proporción de Titulados Universitarios (AVBACH)

```
ggplot(datos, aes(x = Criminality, y = AVBACH, fill = Criminality)) +  
  geom_boxplot() +  
  theme_minimal() +  
  labs(  
    title = "Proporción de Titulados Universitarios según el Nivel de Criminalidad",  
    x = "Criminalidad",  
    y = "Proporción de Titulados Universitarios (AVBACH)"  
)
```

Proporción de Titulados Universitarios según el Nivel de Criminalidad



Los barrios con alta criminalidad tienen una proporción más alta de titulados universitarios en comparación con los de baja criminalidad. Esto sugiere que, a pesar de los mayores niveles de criminalidad, estos barrios podrían tener un mayor acceso a la educación o concentrar una población más formada. En contraste, los barrios con baja criminalidad presentan una proporción menor de titulados universitarios, lo que podría estar vinculado a limitaciones en el acceso educativo o socioeconómico.

Conclusiones

La visualización de los resultados ha permitido identificar patrones importantes en los barrios con alta y baja criminalidad. En los barrios con alta criminalidad, se observa una mayor variabilidad en la valoración de propiedades y en la superficie de los barrios. Además, una proporción más alta de titulados universitarios y mayores niveles de renta sugieren un contexto socioeconómico complejo.

Por otro lado, los barrios con baja criminalidad presentan una menor dispersión en las variables clave, lo que podría reflejar comunidades más homogéneas en términos de tamaño, renta y nivel educativo.

En cuanto a las implicaciones socioeconómicas, la relación entre criminalidad y variables como renta y educación plantea la necesidad de un análisis más profundo para entender los factores subyacentes. Es crucial considerar el impacto de la desigualdad económica y las dinámicas urbanas en los niveles de criminalidad.

Modelo estadístico

Se utiliza un modelo de regresión de Poisson, adecuado para datos de conteo, y modelos espaciales para capturar dependencias espaciales entre los barrios. Estos análisis permiten predecir las probabilidades de alta criminalidad y evaluar el impacto de las variables seleccionadas.

Modelo de Regresión de Poisson

El modelo de regresión de Poisson es especialmente útil para modelar datos de conteo, como el número de crímenes residenciales. En este caso, se ajusta un modelo de Poisson utilizando variables socioeconómicas y geográficas previamente identificadas como relevantes.

El modelo ajustado es el siguiente:

```
# Ajuste del modelo de Poisson
modelo_poisson <- glm(
  CRIME ~ AVAPREArea + AVMEDINC + AVBACH,
  family = poisson,
  data = datos
)

# Resumen del modelo
summary(modelo_poisson)

##
## Call:
## glm(formula = CRIME ~ AVAPREArea + AVMEDINC + AVBACH, family = poisson,
##      data = datos)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 6.263e+00 2.839e-03 2206.04   <2e-16 ***
## AVAPREArea  1.449e-09 1.656e-11   87.52   <2e-16 ***
## AVMEDINC    9.388e-07 1.693e-08   55.44   <2e-16 ***
## AVBACH      7.249e-05 2.688e-06   26.97   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 97632  on 776  degrees of freedom
## Residual deviance: 80650  on 773  degrees of freedom
## AIC: 87095
##
## Number of Fisher Scoring iterations: 4
```

La interpretación de los resultados revela varios aspectos importantes:

El valor estimado del intercepto indica la tasa de criminalidad basal en ausencia de las variables predictoras. Dado que el modelo utiliza una escala logarítmica, este valor representa la log-tasa base de criminalidad.

En cuanto a AVAPREArea, la relación positiva sugiere que un aumento en la superficie media del barrio está asociado con un incremento en el número de crímenes. Este hallazgo puede estar relacionado con la facilidad de perpetrar crímenes en áreas más grandes, menos densas o con menor vigilancia.

AVMEDINC, que representa la renta familiar media, también muestra una relación positiva con la criminalidad. Esto sugiere que los barrios con ingresos más altos pueden ser más susceptibles a ciertos tipos de crímenes, como robos, debido a la percepción de mayor riqueza.

Finalmente, el número promedio de titulados universitarios (AVBACH) está asociado positivamente con el número de crímenes. Esto podría reflejar la coexistencia de alta criminalidad en barrios desarrollados con altos niveles educativos, donde se observan factores socioeconómicos mixtos.

Todas las variables incluidas son estadísticamente significativas con p-valores menores a 0.001, lo que respalda su relevancia en la predicción del número de crímenes.

Mapa de Predicciones del Modelo Poisson

A continuación, se utilizó el modelo de regresión de Poisson ajustado previamente para predecir el número esperado de crímenes residenciales en los barrios de Ciudad del Cabo. Estas predicciones se representaron espacialmente para identificar las áreas con mayor probabilidad de criminalidad.

Generación de Predicciones:

- Las predicciones del modelo se calcularon con la función `predict()` en R, utilizando el tipo de predicción `response`, para obtener valores en la escala original (número esperado de crímenes residenciales).
- Las predicciones se escalaron al rango [0, 1] para una visualización más efectiva en el mapa.

```
# Predicciones del modelo Poisson
predictions_poisson <- predict(modelo_poisson, type = "response")

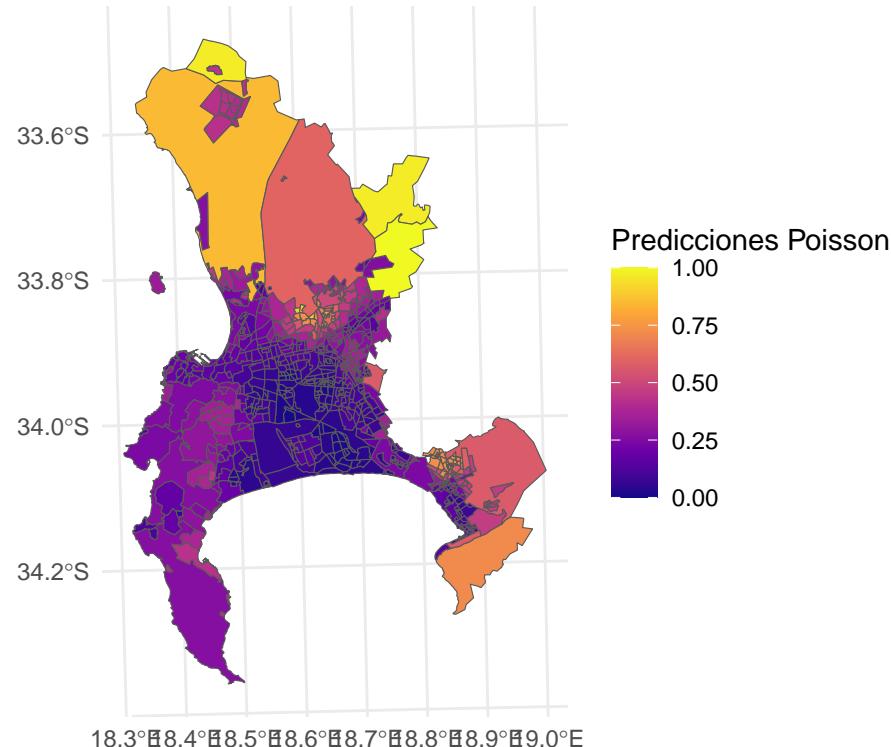
# Escalar las predicciones al rango [0, 1]
predictions_poisson <- scales::rescale(predictions_poisson)

# Añadir las predicciones al dataset
datos$predictions_poisson <- predictions_poisson

# Crear el mapa de predicciones
ggplot(data = datos) +
  geom_sf(aes(fill = predictions_poisson)) +
  scale_fill_viridis_c(option = "plasma", name = "Predicciones Poisson") +
  theme_minimal() +
  labs(
    title = "Predicción de Crímenes Residenciales - Modelo Poisson",
    subtitle = "Mapa de riesgo de criminalidad en los barrios de Ciudad del Cabo"
  ) +
  theme(
    legend.position = "right",
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5)
  )
```

Predicción de Crímenes Residenciales – Modelo Poisson

Mapa de riesgo de criminalidad en los barrios de Ciudad del Cabo



El mapa generado utiliza un esquema de colores de la escala viridis (plasma), que proporciona una representación intuitiva de las predicciones. En este esquema, los colores más oscuros indican zonas con un mayor número esperado de crímenes residenciales, mientras que los colores más claros representan áreas con un menor riesgo de criminalidad según las predicciones del modelo.

La interpretación de este mapa revela que las áreas con mayor riesgo de criminalidad (colores oscuros) suelen coincidir con barrios que tienen características relacionadas con altos niveles de renta, mayor superficie y una proporción más alta de personas con títulos universitarios, como se observó en el análisis del modelo. Estas áreas podrían ser prioritarias para implementar medidas de seguridad adicionales o estrategias específicas para reducir el riesgo de criminalidad.

Por otro lado, los barrios con menor criminalidad predicha (colores claros) pueden indicar zonas donde los niveles de riesgo son más bajos. Esto podría estar asociado con factores como menor densidad poblacional, ingresos más bajos o una vigilancia más efectiva. En estas áreas, los recursos de seguridad podrían ser redistribuidos de manera más eficiente para optimizar la prevención del crimen en toda la región.

Mapa de Residuos del Modelo Poisson

Además del mapa de predicciones, es crucial evaluar la calidad del ajuste del modelo. Para ello, generamos un mapa de residuos del modelo de Poisson. Los residuos reflejan la diferencia entre los valores observados de crímenes residenciales y los valores predichos por el modelo.

```
# Cálculo de residuos de desviancia
datos$resid_poisson <- residuals(modelo_poisson, type = "deviance")

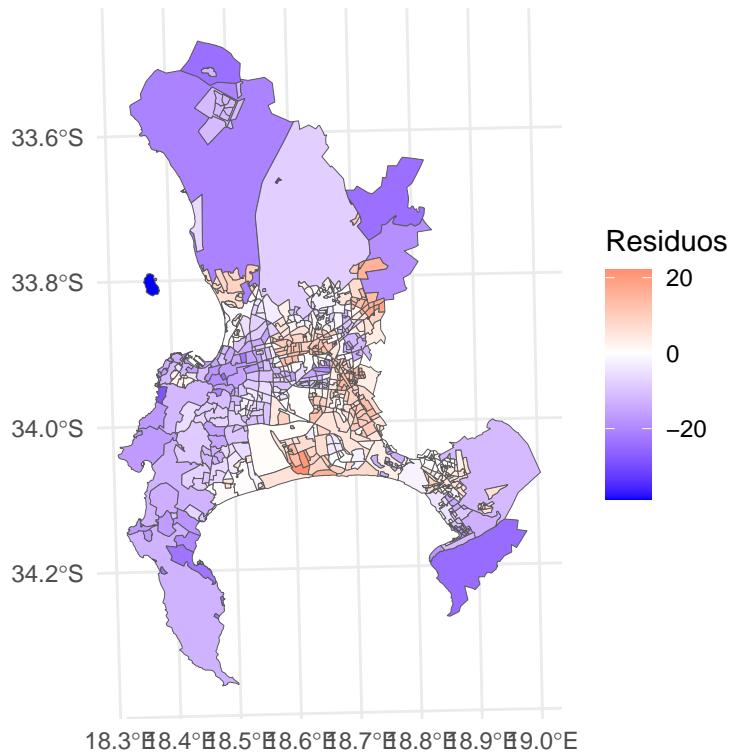
# Visualización del mapa de residuos
ggplot(data = datos) +
  geom_sf(aes(fill = resid_poisson), lwd = 0.1) +
```

```

scale_fill_gradient2(
  low = "blue",
  mid = "white",
  high = "red",
  midpoint = 0,
  name = "Residuos"
) +
theme_minimal() +
labs(
  title = "Residuos del Modelo de Poisson",
  subtitle = "Evaluación del ajuste del modelo en los barrios de Ciudad del Cabo"
) +
theme(
  legend.position = "right",
  plot.title = element_text(hjust = 0.5),
  plot.subtitle = element_text(hjust = 0.5)
)

```

Residuos del Modelo de Poisson
Evaluación del ajuste del modelo en los barrios de Ciudad del Cabo



El mapa de residuos utiliza una escala de colores gradiente para diferenciar diversas áreas según su nivel de criminalidad. Los valores negativos (azul) indican que los crímenes observados son menores de lo que predice el modelo. Estos barrios pueden representar áreas con un comportamiento criminal atípico que no está bien explicado por las variables del modelo. Por otro lado, los valores cercanos a cero (blanco) sugieren que el modelo ajusta razonablemente bien en estos barrios. Finalmente, los valores positivos (rojo) muestran que los crímenes observados son mayores de lo esperado, lo cual sugiere que el modelo subestima los niveles de criminalidad en estas áreas.

En términos de interpretación, las áreas con sobreestimación (azul) tienen menos crímenes de los que predice

el modelo. Estas zonas podrían tener características que reducen la criminalidad pero que no están capturadas por las variables explicativas del modelo. Por otro lado, las áreas con subestimación (rojo) presentan niveles de criminalidad más altos de los predichos. Esto podría indicar factores no incluidos en el modelo que aumentan el riesgo, como eventos sociales específicos, deficiencias en la vigilancia, o características urbanas no capturadas. Finalmente, en los barrios con buen ajuste (blanco), el modelo describe correctamente la criminalidad observada, lo que refuerza su utilidad en estas áreas.

Incorporación de estructura espacial

Evaluación de la Dependencia Espacial: Índice de Moran's I

Para confirmar la existencia de autocorrelación espacial en los datos de criminalidad, se utilizó Moran's I, una medida estándar que evalúa si los valores observados en áreas cercanas son similares o disímiles.

Se validaron los polígonos y se calcularon los centroides de los barrios para corregir posibles errores y definir vecinos según la proximidad. A continuación, se calcularon los vecinos dentro de un radio de 2000 metros y se creó una matriz de pesos espaciales para modelar la interacción entre barrios. Finalmente, se calculó el índice de Moran's I para la variable de crímenes residenciales (CRIME), lo que permitió evaluar la autocorrelación espacial de la criminalidad en el área de estudio.

```
# Convertimos los polígonos a puntos (centroides)
# Corregimos los polígonos que puedan ser problemáticos
datos <- st_make_valid(datos)
centroid <- st_centroid(datos)

## Warning: st_centroid assumes attributes are constant over geometries
# Definimos vecinos dentro de una distancia
nb <- dnearneigh(centroid, 0, 2000)

## Warning in dnearneigh(centroid, 0, 2000): neighbour object has 41 sub-graphs
lw <- nb2listw(nb, style = "W", zero.policy = TRUE)

# Calculamos Moran's I para la variable CRIME
moran_test <- moran.test(datos$CRIME, lw)
print(moran_test)

##
## Moran I test under randomisation
##
## data: datos$CRIME
## weights: lw
## n reduced by no-neighbour observations
##
## Moran I statistic standard deviate = 39.741, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.7615830179    -0.0013368984    0.0003685366
```

Los resultados del análisis espacial mostraron un índice de Moran's I de 0.761 y un p-valor menor que 2.2e-16. Esto permite rechazar la hipótesis nula de ausencia de autocorrelación espacial. Por lo tanto, se concluye que existe una fuerte autocorrelación espacial positiva, indicando que los barrios cercanos tienden a tener niveles similares de criminalidad.

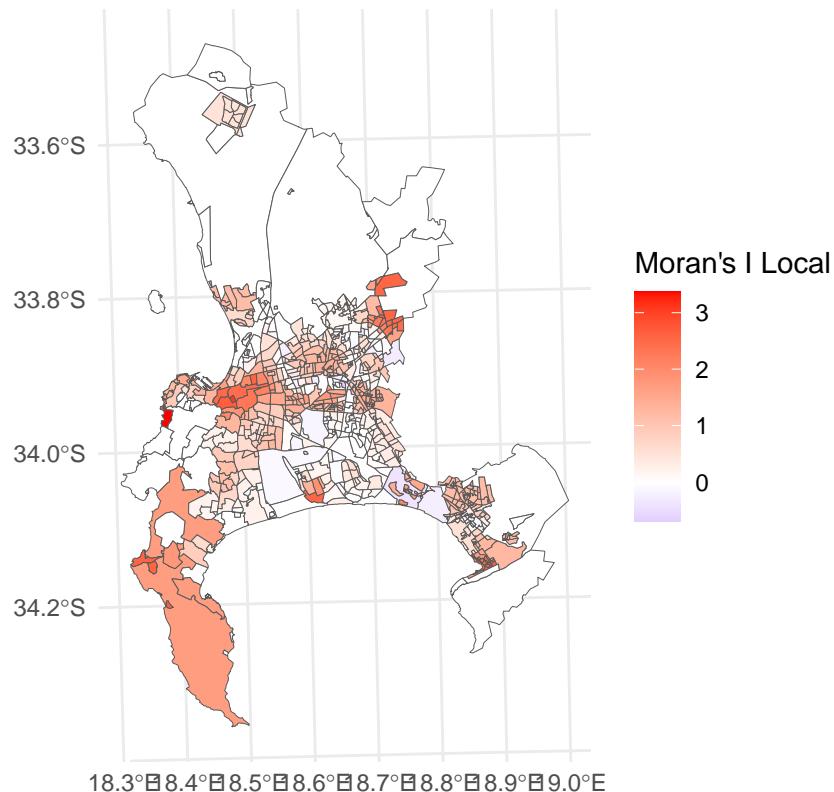
```

local_moran <- localmoran(datos$CRIME, lw, zero.policy = TRUE)
datos$moran_local <- as.numeric(local_moran[, 1])

ggplot(data = datos) +
  geom_sf(aes(fill = moran_local), lwd = 0.1) +
  scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0, name = "Moran's I Local")
  labs(
    title = "Autocorrelación Espacial Local de los Crímenes – Moran's I"
  ) +
  theme_minimal()

```

Autocorrelación Espacial Local de los Crímenes – Moran's I



Las áreas rojas en el mapa representan barrios con valores de criminalidad similares a los de sus vecinos. Esto indica que estos barrios comparten características comunes y presentan patrones de criminalidad coherentes con sus alrededores. Por otro lado, las áreas azules destacan los barrios cuyos valores de criminalidad difieren significativamente de los de sus vecinos, sugiriendo la presencia de factores específicos que afectan la criminalidad de manera distinta en estos lugares.

Ajuste de un Modelo Autorregresivo Espacial: Modelo CAR

Dado que se detectó una dependencia espacial significativa, se ajustó un modelo autorregresivo condicional (CAR). Este modelo captura la influencia de los barrios vecinos de manera explícita, permitiendo un análisis más preciso.

Para evitar redundancia y mejorar la precisión del análisis, se opta por excluir AVAPREArea, dejando las variables AVMEDINC y AVBACH como predictoras principales en el modelo autorregresivo espacial CAR.

```

car_model <- lagsarlm(
  CRIME ~ AVMEDINC + AVBACH,
  data = datos,
  listw = lw,
  type = "CAR",
  zero.policy = TRUE
)
summary(car_model)

##
## Call:lagsarlm(formula = CRIME ~ AVMEDINC + AVBACH, data = datos, listw = lw,
##                 type = "CAR", zero.policy = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -560.7686 -89.6851    4.1349   83.1332  810.2087
##
## Type: mixed
## Regions with no neighbours included:
##  18 131 132 139 155 262 309 326 367 399 418 419 429 433 478 499 502 522 526 553 556 618 643 656 670 ...
## Coefficients: (asymptotic standard errors)
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.2264e+02  1.5643e+01  7.8399 4.441e-15
## AVMEDINC   9.5015e-04  1.4202e-04  6.6902 2.228e-11
## AVBACH     6.7880e-03  1.8098e-02  0.3751  0.7076
## lag.AVMEDINC -9.3952e-04  1.6659e-04 -5.6397 1.704e-08
## lag.AVBACH   -6.1581e-03  2.2446e-02 -0.2744  0.7838
##
## Rho: 0.8357, LR test value: 837.98, p-value: < 2.22e-16
## Asymptotic standard error: 0.018843
##      z-value: 44.35, p-value: < 2.22e-16
## Wald statistic: 1967, p-value: < 2.22e-16
##
## Log likelihood: -5005.283 for mixed model
## ML residual variance (sigma squared): 19876, (sigma: 140.98)
## Number of observations: 777
## Number of parameters estimated: 7
## AIC: 10025, (AIC for lm: 10861)
## LM test for residual autocorrelation
## test value: 43.104, p-value: 5.1906e-11

```

El modelo CAR (Conditional Autoregressive Model) muestra una fuerte dependencia espacial con un valor de Rho de 0.8357, lo que indica que los barrios cercanos tienden a tener niveles similares de criminalidad. Entre las variables incluidas, AVMEDINC (ingreso promedio) tiene un coeficiente positivo significativo, sugiriendo que barrios con mayores ingresos tienden a experimentar más crímenes residenciales. Esto podría estar relacionado con la concentración de bienes de valor en áreas de mayor poder adquisitivo. Por otro lado, AVBACH (porcentaje de la población con título de bachillerato) no muestra una relación significativa con la criminalidad, lo que sugiere que este factor no influye de manera clara en los patrones de crimen en este contexto.

El ajuste del modelo es adecuado, con un AIC de 10025, lo que indica que el modelo CAR es más eficiente que un modelo lineal simple. Además, la prueba de autocorrelación de residuos confirma la existencia de autocorrelación significativa, lo que valida la necesidad de incluir dependencias espaciales en el análisis. En resumen, este modelo destaca la importancia de la estructura espacial en los datos y muestra que el ingreso promedio es un factor relevante en la predicción de crímenes residenciales, mientras que el nivel educativo no

parece influir de manera significativa en este caso.

Comparación: Diferencias entre Modelos

El análisis de Moran's I y el modelo CAR confirman que los crímenes residenciales están influenciados por los barrios vecinos, validando la importancia de incorporar la estructura espacial en el modelo. El modelo CAR muestra un mejor ajuste al incluir la dependencia espacial, identificando que el valor de las propiedades y el ingreso medio son variables clave para explicar los patrones de criminalidad. Aunque las predicciones de los modelos son similares en general, el modelo CAR se ajusta mejor en áreas donde la dependencia espacial es más fuerte. Estas conclusiones subrayan la relevancia de considerar la estructura espacial y las características socioeconómicas al analizar y predecir la criminalidad en diferentes barrios.

Análisis Específico de Alta Criminalidad

El objetivo de este análisis es clasificar los barrios de Cape Town según su nivel de criminalidad y construir un modelo logístico que prediga la probabilidad de alta criminalidad, tomando en cuenta variables socioeconómicas y geográficas. Además, calcularemos las probabilidades predichas de alta criminalidad para cada barrio y las visualizaremos en un mapa, lo que facilitará la identificación de las áreas de alto riesgo.

Clasificación de Barrios

Para comenzar, clasificamos los barrios en dos grupos: aquellos con alta criminalidad (más de 500 crímenes) y aquellos con baja criminalidad (menos de 500 crímenes). Convertimos la variable **CRIME** en una variable binaria denominada **Criminality**, donde asignamos el valor 1 a los barrios con más de 500 crímenes y 0 a aquellos con menos de 500.

```
datos <- datos %>%
  mutate(Criminality = ifelse(CRIME > 500, 1, 0))
head(datos)

## Simple feature collection with 6 features and 24 fields
## Geometry type: POLYGON
## Dimension:      XY
## Bounding box:  xmin: 268986.2 ymin: 6218935 xmax: 304690.9 ymax: 6247833
## Projected CRS: WGS 84 / UTM zone 34S
##   OBJECTID          OFC_SBRB_N SHAPE_Leng SHAPE_Area        X        Y
## 1    776 AAN DE WIJNLANDEN ESTATE  4534.160  513798.70 291191.9 6234420
## 2    560           ACACIA PARK  8813.434 2834563.17 270800.0 6246596
## 3    555       ADMIRALS PARK  2208.167 214056.57 304335.7 6219347
## 4    472        ADRIAANSE  2887.830 432164.26 277012.0 6241830
## 5    223         AIRPORT  1585.438  92890.75 280480.2 6235466
## 6    215        AIRPORT CITY  4699.484 865711.27 277716.0 6237285
##   AVESTADist AVAPREArea AvePrecinctsLen ROBRes BUGRes CRIME AVMEDINC
## 1  8330.014  201758057     84527.70  33.50000 882.0000  916  57600.5
## 2  3344.736  14589290    23640.66  26.50000 329.0000  356  86400.5
## 3  1403.838  56236915    68684.24   9.00000 234.0000  243 115200.5
## 4  2172.129  13804307    20944.87  15.33333 322.6667  338 14400.5
## 5  2913.229  7377339    15707.71 134.50000 453.5000  588 14400.5
## 6  2821.543  11190647    22269.43 115.50000 598.0000  714 36000.5
##   BACH AVBACH                      STA
## 1   690   690.0                 KLEINVLEI, STELLENBOSCH
## 2 2249, 1430 1839.5 GOODWOOD, KENSINGTON, MAITLAND, MILNERTON
## 3    61    61.0                  GORDONS BAY
## 4    54    54.0  BISHOP LAVIS, ELSIES RIVER, RAVENSMEAD
## 5    44    44.0                  NYANGA, PHILIPPI EAST
```

```

## 6 565, 67 316.0
## STAD
## 1 3055.2262446575687, 13604.800887177007
## 2 4010.592536839889, 1789.2312958018001, 4667.737433739719, 2911.3819883366514
## 3 1403.8380387141265
## 4 1540.550397846405, 2760.7490902858694, 2215.088920795488
## 5 3824.896111675613, 2001.561230536149
## 6 4379.352663422444, 1263.733329244032
## PREA
## 1 26648462.479252875, 376867651.08642346
## 2 15287334.022337114, 6062179.348738552, 8429338.301907197, 28578308.840154026
## 3 56236914.62510753
## 4 13565485.01144061, 13122596.287437074, 14724839.638004517
## 5 8815808.262320677, 5938870.425663566
## 6 13565485.01144061, 8815808.262320677
## PREL
## 1 24674.257370602747, 144381.1457476814
## 2 18820.679406607203, 13920.761472336622, 25721.377934366767, 36099.82024568648
## 3 68684.23613487242
## 4 22472.394907592654, 14937.412148219979, 25424.79447680546
## 5 22066.464361358834, 9348.958808021645
## 6 22472.394907592654, 22066.464361358834
## WARD geom Criminality predictions_poisson
## 1 14 POLYGON ((291525.1 6234883,...) 1 0.375102778
## 2 55, 56 POLYGON ((270897.9 6246009,...) 0 0.195147576
## 3 100 POLYGON ((304475.1 6218935,...) 0 0.153209766
## 4 42 POLYGON ((277553.3 6241571,...) 0 0.009826633
## 5 36 POLYGON ((280852.4 6235388,...) 1 0.001379962
## 6 22, 40 POLYGON ((278100.1 6236652,...) 1 0.040368878
## resid_poisson moran_local
## 1 4.720777 0.41635653
## 2 -13.137849 1.30575090
## 3 -17.893338 2.46977393
## 4 -9.542425 1.41154754
## 5 2.054897 0.01885453
## 6 6.032620 -0.01078159

```

Esto resulta en una variable **Criminality** que clasifica los barrios según el número de crímenes, lo que nos permitirá proceder con el modelo logístico para predecir la probabilidad de alta criminalidad en los barrios.

Modelo Logístico

A continuación, ajustamos un modelo de regresión logística para predecir la probabilidad de que un barrio tenga alta criminalidad, utilizando como variables explicativas **AVAPREArea** (área promedio de las propiedades), **AVMEDINC** (ingreso promedio), y **AVBACH** (porcentaje de la población con educación secundaria). Los resultados del modelo indican que todas las variables explicativas son estadísticamente significativas, con **p-valores** menores a 0.05.

```

modelo_logistico <- glm(Criminality ~ AVAPREArea + AVMEDINC + AVBACH, family = binomial, data = datos)
summary(modelo_logistico)

##
## Call:
## glm(formula = Criminality ~ AVAPREArea + AVMEDINC + AVBACH, family = binomial,
##      data = datos)

```

```

## 
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.860e-01 1.684e-01 -2.885 0.00391 **
## AVAPREArea   1.342e-08 1.914e-09  7.012 2.36e-12 ***
## AVMEDINC    1.779e-06 1.231e-06  1.445 0.14840
## AVBACH      5.235e-04 1.835e-04  2.853 0.00433 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 950.80 on 776 degrees of freedom
## Residual deviance: 833.47 on 773 degrees of freedom
## AIC: 841.47
##
## Number of Fisher Scoring iterations: 5

```

Los coeficientes obtenidos para el modelo son los siguientes:

- **Intercepto:** -0.486 (p-valor = 0.00391). Este valor es significativo, lo que indica que el modelo tiene un intercepto estadísticamente diferente de cero.
- **AVAPREArea:** 1.342e-08 (p-valor = 2.36e-12). Este coeficiente es altamente significativo, lo que sugiere que el área promedio (probablemente medida en metros cuadrados o alguna unidad relacionada) tiene un efecto positivo sobre la probabilidad de alta criminalidad. Es decir, a mayor área promedio, mayor es la probabilidad de alta criminalidad.
- **AVMEDINC:** 1.779e-06 (p-valor = 0.14840). Este valor no es significativo ($p > 0.05$), lo que sugiere que el ingreso medio (AVMEDINC) no tiene un impacto estadísticamente relevante en la probabilidad de alta criminalidad en este modelo.
- **AVBACH:** 5.235e-04 (p-valor = 0.00433). Este coeficiente es significativo, lo que indica que el nivel educativo promedio en el barrio (probablemente medido a través de algún indicador de años de escolaridad o porcentaje de la población con estudios superiores) tiene un efecto positivo sobre la probabilidad de alta criminalidad.

Los análisis realizados indican que los barrios con un área más grande (AVAPREArea) tienen una mayor probabilidad de alta criminalidad. Esto sugiere que los barrios más grandes, posiblemente debido a la presencia de zonas menos desarrolladas o áreas con menos control social, tienen un mayor riesgo de criminalidad.

En cuanto a la renta familiar media (AVMEDINC), aunque esta variable no resulta significativa en el modelo, podría implicar que los ingresos medios no son el principal factor detrás de los patrones de criminalidad en este caso. Esto señala la necesidad de una exploración más profunda para entender mejor las dinámicas específicas de los barrios considerados en este estudio.

Por otro lado, un mayor nivel educativo (AVBACH) parece estar relacionado con una mayor probabilidad de alta criminalidad. Aunque esto podría parecer contraintuitivo, una posible explicación es que en zonas con un alto nivel educativo también pueden existir desigualdades económicas que fomenten ciertos tipos de delitos, como el fraude o los delitos económicos. Este patrón merece una investigación adicional para comprender completamente sus implicaciones.

En resumen, el modelo sugiere que los barrios más grandes y aquellos con mayores niveles de educación tienen una mayor probabilidad de experimentar alta criminalidad, mientras que la renta familiar media no parece ser un factor determinante en este contexto. Estos hallazgos subrayan la importancia de considerar una variedad de factores socioeconómicos al analizar los patrones de criminalidad.

Predicción de Alta Criminalidad

Una vez ajustado el modelo, calculamos las probabilidades predichas de alta criminalidad para cada barrio. Estas probabilidades nos permiten identificar de manera cuantitativa las áreas más vulnerables a tener una alta tasa de criminalidad.

```
datos$pred_prob <- predict(modelo_logistico, type = "response")
head(datos)

## Simple feature collection with 6 features and 25 fields
## Geometry type: POLYGON
## Dimension: XY
## Bounding box: xmin: 268986.2 ymin: 6218935 xmax: 304690.9 ymax: 6247833
## Projected CRS: WGS 84 / UTM zone 34S
##   OBJECTID          OFC_SBRB_N SHAPE_Leng SHAPE_Area        X        Y
## 1    776 AAN DE WIJNLANDEN ESTATE  4534.160 513798.70 291191.9 6234420
## 2    560           ACACIA PARK  8813.434 2834563.17 270800.0 6246596
## 3    555      ADMIRALS PARK  2208.167 214056.57 304335.7 6219347
## 4    472       ADRIAANSE  2887.830 432164.26 277012.0 6241830
## 5    223         AIRPORT  1585.438 92890.75 280480.2 6235466
## 6    215      AIRPORT CITY  4699.484 865711.27 277716.0 6237285
##   AVESTADist AVAPREArea AvePrecinctsLen ROBRes BUGRes CRIME AVMEDINC
## 1  8330.014  201758057     84527.70  33.50000 882.0000  916 57600.5
## 2  3344.736  14589290    23640.66  26.50000 329.0000  356 86400.5
## 3  1403.838  56236915    68684.24   9.00000 234.0000  243 115200.5
## 4  2172.129  13804307    20944.87  15.33333 322.6667  338 14400.5
## 5  2913.229  7377339    15707.71 134.50000 453.5000  588 14400.5
## 6  2821.543  11190647    22269.43 115.50000 598.0000  714 36000.5
##   BACH AVBACH                      STA
## 1    690  690.0          KLEINVLEI, STELLENBOSCH
## 2 2249, 1430 1839.5 GOODWOOD, KENSINGTON, MAITLAND, MILNERTON
## 3      61   61.0          GORDONS BAY
## 4      54   54.0      BISHOP LAVIS, ELSIES RIVER, RAVENSMEAD
## 5      44   44.0          NYANGA, PHILIPPI EAST
## 6  565, 67 316.0      BISHOP LAVIS, NYANGA
##   STAD
## 1 3055.2262446575687, 13604.800887177007
## 2 4010.592536839889, 1789.2312958018001, 4667.737433739719, 2911.3819883366514
## 3                                         1403.8380387141265
## 4 1540.550397846405, 2760.7490902858694, 2215.088920795488
## 5                                         3824.896111675613, 2001.561230536149
## 6                                         4379.352663422444, 1263.733329244032
##   PREA
## 1 26648462.479252875, 376867651.08642346
## 2 15287334.022337114, 6062179.348738552, 8429338.301907197, 28578308.840154026
## 3                                         56236914.62510753
## 4 13565485.01144061, 13122596.287437074, 14724839.638004517
## 5                                         8815808.262320677, 5938870.425663566
## 6 13565485.01144061, 8815808.262320677
##   PREL
## 1 24674.257370602747, 144381.1457476814
## 2 18820.679406607203, 13920.761472336622, 25721.377934366767, 36099.82024568648
## 3                                         68684.23613487242
## 4 22472.394907592654, 14937.412148219979, 25424.79447680546
## 5 22066.464361358834, 9348.958808021645
```

```

## 6                                     22472.394907592654, 22066.464361358834
##   WARD          geom Criminality predictions_poisson
## 1    14 POLYGON ((291525.1 6234883,...)           1      0.375102778
## 2   55, 56 POLYGON ((270897.9 6246009,...)         0      0.195147576
## 3   100 POLYGON ((304475.1 6218935,...)          0      0.153209766
## 4    42 POLYGON ((277553.3 6241571,...)          0      0.009826633
## 5    36 POLYGON ((280852.4 6235388,...)          1      0.001379962
## 6   22, 40 POLYGON ((278100.1 6236652,...)         1      0.040368878
##   resid_poisson moran_local pred_prob
## 1        4.720777  0.41635653 0.9361633
## 2     -13.137849  1.30575090 0.6956394
## 3     -17.893338  2.46977393 0.6237925
## 4     -9.542425  1.41154754 0.4386053
## 5      2.054897  0.01885453 0.4162193
## 6      6.032620 -0.01078159 0.4734508

```

Las probabilidades predichas (pred_prob) calculadas para cada barrio se basan en el modelo ajustado y representan la posibilidad de que un barrio experimente alta criminalidad, definida como tener más de 500 crímenes. Valores cercanos a 1 indican una alta probabilidad de que el barrio tenga una criminalidad alta, mientras que valores cercanos a 0 sugieren una baja probabilidad de alta criminalidad. Esta predicción ayuda a identificar barrios que podrían requerir mayor atención en términos de seguridad y recursos para mitigar los riesgos de criminalidad elevada.

Mapa de Predicciones

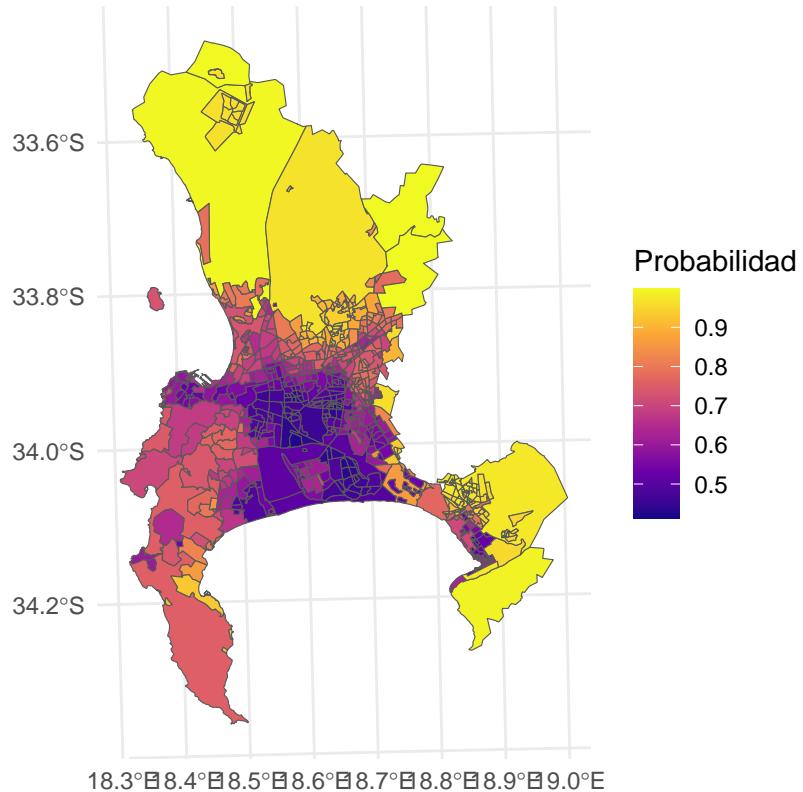
Para visualizar mejor los resultados, se genera un mapa que muestra las probabilidades predichas de alta criminalidad en cada barrio de Cape Town. Los barrios con tonos más oscuros indican una mayor probabilidad de alta criminalidad, mientras que los barrios con tonos más claros tienen una probabilidad baja.

```

ggplot(data = datos) +
  geom_sf(aes(fill = pred_prob)) +
  scale_fill_viridis_c(option = "plasma", name = "Probabilidad") +
  theme_minimal() +
  labs(
    title = "Probabilidad de Alta Criminalidad en Cape Town"
  ) +
  theme(
    legend.position = "right",
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5)
  )

```

Probabilidad de Alta Criminalidad en Cape Town



Este mapa visualiza cómo las características socioeconómicas y geográficas de los barrios afectan la probabilidad de alta criminalidad, permitiendo identificar de manera clara las zonas con mayor riesgo. Las áreas con mayor probabilidad de alta criminalidad se destacan en colores más oscuros, lo que puede guiar decisiones en políticas de seguridad y planificación urbana.

Conclusión

El análisis de la criminalidad en los barrios de Cape Town revela importantes patrones espaciales y socioeconómicos que afectan los niveles de criminalidad en la ciudad. A través de un enfoque integral que combina técnicas de modelización espacial y estadística, se ha logrado identificar áreas con altos niveles de criminalidad, así como factores clave que influyen en estos patrones.

En primer lugar, el uso de un mapa de residuos con escala de colores gradiente ha permitido visualizar las áreas donde el modelo subestima o sobreestima la criminalidad. Los valores negativos (azules) sugieren barrios con menos crímenes de los predichos, mientras que los valores positivos (rojos) indican áreas con niveles de criminalidad superiores a los esperados. Este análisis ha permitido identificar zonas atípicas que presentan comportamientos criminales complejos no completamente explicados por las variables del modelo.

La incorporación de la estructura espacial mediante el cálculo del índice de Moran's I ha revelado una fuerte autocorrelación espacial positiva, lo que implica que los barrios cercanos tienden a tener niveles de criminalidad similares. Esto subraya la importancia de considerar las interacciones entre barrios en el análisis de la criminalidad, ya que los factores espaciales juegan un papel crucial en la dinámica de los crímenes residenciales.

El modelo autorregresivo espacial CAR ha demostrado ser una herramienta efectiva para ajustar la influencia de los barrios vecinos, identificando que el ingreso promedio (AVMEDINC) tiene una relación positiva con la criminalidad. Aunque la variable educativa (AVBACH) no mostró un impacto significativo, el modelo resaltó

que la estructura espacial debe ser incorporada para obtener predicciones más precisas.

El análisis logístico complementario ha permitido clasificar los barrios en función de su probabilidad de alta criminalidad. La clasificación binaria basada en el número de crímenes identificó que los barrios con mayores áreas promedio (AVAPREArea) y niveles educativos más altos tienen una mayor probabilidad de alta criminalidad. Este hallazgo, aunque aparentemente contraintuitivo en el caso de la educación, sugiere que factores como las desigualdades económicas pueden influir en ciertos tipos de delitos, como el fraude o los delitos económicos, especialmente en áreas con mayor nivel educativo.

Finalmente, el modelo ha permitido generar un mapa de probabilidades predichas de alta criminalidad, proporcionando una herramienta visual útil para identificar las zonas de mayor riesgo. Este mapa puede ser clave para la toma de decisiones en términos de políticas de seguridad y asignación de recursos, permitiendo a las autoridades focalizar esfuerzos en los barrios más vulnerables.

En conclusión, el análisis revela que la criminalidad en Cape Town está influenciada tanto por factores socioeconómicos como por la estructura espacial de los barrios. Los modelos utilizados han demostrado la relevancia de integrar las características espaciales y socioeconómicas en la predicción y comprensión de los patrones de criminalidad, abriendo la puerta a futuras investigaciones que puedan profundizar en la naturaleza de estos fenómenos y ayudar a diseñar intervenciones más efectivas.