

Autocorrelação Espacial do Crime de Los Angeles

Erivando Sena (Com adaptações do original de Ambarish Ganguly)

22 de maio de 2019

O objetivo desta versão adaptada utilizando R Notebook é reapresentar à análise realizada no idioma Português Brasileiro para que mais pessoas possam obter acesso ao conteúdo publicado originalmente em inglês na Comunidade Kaggle.

Introdução

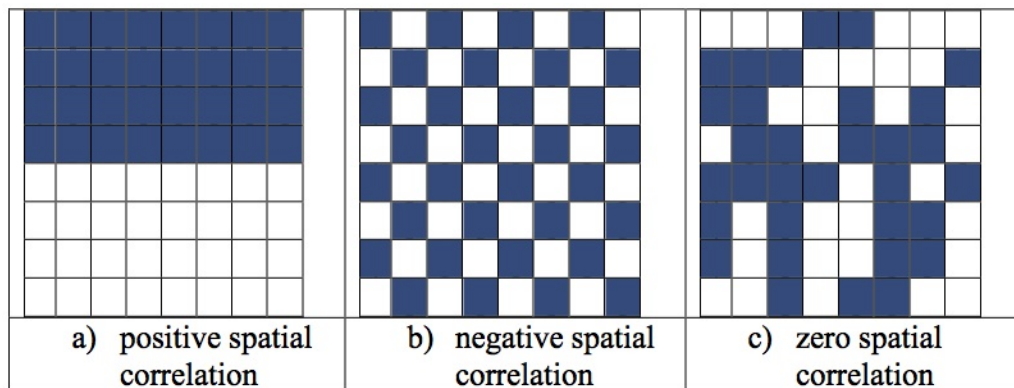
Esta análise busca investigar um método de cálculo de correlação espacial em R. Consideraremos as estatísticas de autocorrelação espacial global e como identificar agrupamentos espaciais em toda a nossa área de estudo.

A base deste trabalho é a Primeira Lei da Geografia de Tobler: “tudo está relacionado a tudo mais, mas as coisas próximas estão mais relacionadas do que as coisas distantes”.

Assim, usando o conjunto de dados (Dados de Crime Polícia de Los Angeles), veremos se o crime em Los Angeles (Analisado na Comunidade Kaggle) é espacialmente autocorrelacionado e identificamos os clusers usando o LISA (Local Measure of Spatial Association).

Autocorrelação Espacial

```
# Exibindo imagem.  
knitr::include_graphics("./Spatial_Autocorrelation.jpg")
```



Preparando o ambiente

```
# Setando idioma e local da pasta de trabalho  
Sys.setlocale("LC_ALL", "Portuguese_Brazil.1252")  
  
## [1] "LC_COLLATE=Portuguese_Brazil.1252;LC_CTYPE=Portuguese_Brazil.1252;LC_MONETARY=Portuguese_Brazil.1252;LC_NUMERIC=Portuguese_Brazil.1252;LC_TIME=Portuguese_Brazil.1252"   
setwd("E:/Outras Analises R/Crime-em-Los-Angeles")  
  
# Setando codificacao para representar qualquer caractere universal padrao Unicode.  
options(encoding="UTF-8")  
  
# Verificar e instalar o(s) pacote(s) a utilizar na analise.  
pacotes_analise <- c('tidyverse', 'classInt', 'mapttools', 'rgdal', 'RColorBrewer', 'spdep', 'tmap')  
# Necessarios apenas para RMarkdown.  
pacotes_padrao <- c("installr", "rmarkdown", "tinytex", "prettydoc")
```

```

if (length(setdiff(c(pacotes_padrao, pacotes_analise), rownames(installed.packages())) > 0) {
  install.packages(setdiff(c(pacotes_padrao, pacotes_analise), rownames(installed.packages())))
  # RMarkdown precisa de Pandoc e MiKTeX instalados. https://miktex.org/2.9/setup.
  install.pandoc()
}

# Importando bibliotecas.
library(tidyverse)
library(classInt)
library(mapttools)
library(rgdal)
library(tidyr)
library(RColorBrewer)
library(spdep)
library(plyr)
library(tmap)

```

Preparação dos dados

```

# Importando Dados.
districts <- readOGR(dsn = "../LAPD_Reporting_Districts.shp")

```

```

## OGR data source with driver: ESRI Shapefile
## Source: "E:\Outras Analises R\Crime-em-Los-Angeles\Codigo R\LAPD_Reporting_Districts.shp", layer: "LAPD_Reporting_Districts.shp"
## with 1135 features
## It has 7 fields
## Integer64 fields read as strings: OBJECTID REPDIST PREC

```

```

crime = read.csv("../Dados/Crime_Data_2010_2017.csv")

```

```

head(crime,5)

```

```

##   DR.Number Date.Reported Date.Occurred Time.Occurred Area.ID   Area.Name
## 1   1208575   03/14/2013   03/11/2013           1800    12 77th Street
## 2  102005556   01/25/2010   01/22/2010           2300     20   Olympic
## 3     418     03/19/2013   03/18/2013           2030     18   Southeast
## 4  101822289   11/11/2010   11/10/2010           1800     18   Southeast
## 5   42104479   01/11/2014   01/04/2014           2300     21   Topanga
##   Reporting.District Crime.Code      Crime.Code.Description
## 1                1241        626  INTIMATE PARTNER - SIMPLE ASSAULT
## 2                2071        510      VEHICLE - STOLEN
## 3                1823        510      VEHICLE - STOLEN
## 4                1803        510      VEHICLE - STOLEN
## 5                2133       745 VANDALISM - MISDEAMEANOR ($399 OR UNDER)
##   MO.Codes Victim.Age Victim.Sex Victim.Descent Premise.Code
## 1 0416 0446 1243 2000        30      F      W          502
## 2                NA      NA      NA      NA          101
## 3                12      NA      NA      NA          101
## 4                NA      NA      NA      NA          101
## 5          0329        84      M      W          501
##   Premise.Description Weapon.Used.Code
## 1 MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)      400
## 2                STREET      NA
## 3                STREET      NA
## 4                STREET      NA

```

```
## 5          SINGLE FAMILY DWELLING          NA
##          Weapon.Description Status.Code
## 1 STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)      AO
## 2                                                    IC
## 3                                                    IC
## 4                                                    IC
## 5                                                    IC
## Status.Description Crime.Code.1 Crime.Code.2 Crime.Code.3 Crime.Code.4
## 1      Adult Other          626          NA          NA          NA
## 2      Invest Cont          510          NA          NA          NA
## 3      Invest Cont          510          NA          NA          NA
## 4      Invest Cont          510          NA          NA          NA
## 5      Invest Cont          745          NA          NA          NA
##          Address Cross.Street
## 1 6300      BRYNHURST          AV
## 2          VAN NESS          15TH
## 3 200 E 104TH          ST
## 4          88TH          WALL
## 5 7200      CIRRUS          WY
##          Location
## 1 (33.9829, -118.3338)
## 2 (34.0454, -118.3157)
## 3 (33.942, -118.2717)
## 4 (33.9572, -118.2717)
## 5 (34.2009, -118.6369)
```

```
head(districts@data, 5)
```

```
## OBJECTID REPDIST PREC      APREC      BUREAU BASICCAR
## 0      1      1901  19      MISSION VALLEY BUREAU      19A1
## 1      2      1705  17 DEVONSHIRE VALLEY BUREAU      17A35
## 2      3      1902  19      MISSION VALLEY BUREAU      19A7
## 3      4      1903  19      MISSION VALLEY BUREAU      19A7
## 4      5      1904  19      MISSION VALLEY BUREAU      19A7
##          TOOLTIP
## 0      Bureau: VALLEY BUREAU\\nDistrict: 1901\\nDivision: MISSION
## 1 Bureau: VALLEY BUREAU\\nDistrict: 1705\\nDivision: DEVONSHIRE
## 2      Bureau: VALLEY BUREAU\\nDistrict: 1902\\nDivision: MISSION
## 3      Bureau: VALLEY BUREAU\\nDistrict: 1903\\nDivision: MISSION
## 4      Bureau: VALLEY BUREAU\\nDistrict: 1904\\nDivision: MISSION
```

Observando rapidamente os dados, podemos ver que o shapefile é composto de 1135 recursos diferentes. Isto corresponde ao distrito declarante em LA: *crimeReporting.Districtanddistricts@data*REPDIST.

Usando as colunas comuns, podemos tentar mesclar os dados. Primeiro precisamos somar o número de crimes em cada distrito.

```
w <- table(crime$Reporting.District)
rep.dis <- as.data.frame(w)
length(unique(rep.dis$Var1))
```

```
## [1] 1280
```

```
head(rep.dis, 5)
```

```
## Var1 Freq
## 1 100 11
```

```
## 2 101 1119
## 3 105 381
## 4 109 33
## 5 111 3926
```

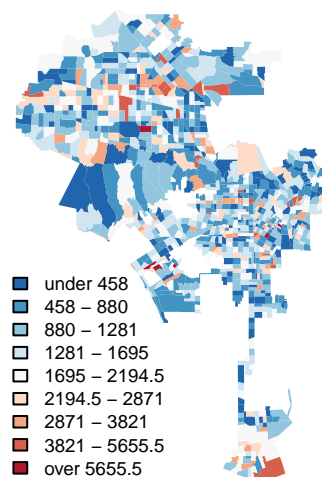
Como podemos ver acima, o número de valores únicos não é igual ao número de recursos no shapefile, isto representa 0 crimes cometidos nesses distritos. Portanto, ao entrar, mantemos essas colunas usando “all.x = TRUE”. Isso significa que mantemos todas as linhas de dados no shapefile. Estes retornam resultados de NA, por padrão, então mudamos isso com a segunda linha de código.

```
districts@data <- merge(districts@data, rep.dis, by.x = "REPDIST", by.y = "Var1", all.x = TRUE)
districts$Freq[is.na(districts$Freq)] <- 0
length(districts$Freq)
```

```
## [1] 1135
```

Plotando Freqüência Criminal

```
var <- districts@data[, "Freq"]
breaks <- classIntervals(var, n = 9, style = "fisher")
my_colours <- rev(brewer.pal(9, "RdBu"))
plot(districts, col = my_colours[findInterval(var, breaks$brks, all.inside = TRUE)],
     axes = FALSE, border = NA)
legend(x = -118.7, y = 34, legend = leglabs(breaks$brks), fill = my_colours, bty = "n", cex = 0.6)
```



Uma rápida visão geral da trama mostra a maioria dos distritos com menores contagens de crimes denunciados. Existem casos esporádicos de altas contagens, especialmente em áreas ao norte, o que poderia sugerir uma autocorrelação espacial próxima a zero, representando a aleatoriedade na distribuição espacial dos dados.

Uma medida melhor seria as taxas de criminalidade, no entanto, a população por distritos reportados não está disponível.

Autocorrelação Espacial

Em vez de especular usando o gráfico, vamos tentar quantificar a correlação. A autocorrelação espacial mede como a distância influencia uma variável específica, neste caso a frequência do crime em um distrito.

Usando a primeira lei de geografia de Tobler, esperamos que a maioria dos fenômenos geográficos exerça uma autocorrelação espacial de algum tipo. Este é frequentemente o caso em dados com fatores humanos, já que pessoas de características similares tendem a residir em vizinhanças semelhantes devido a uma variedade de razões, incluindo preços de casas, proximidade a locais de trabalho e fatores culturais.

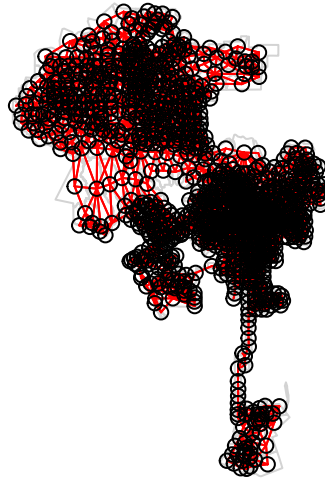
A autocorrelação espacial pode ser representada de duas maneiras: global ou localmente. O modelo global cria uma medida individual representando todo o conjunto de dados, enquanto os modelos locais nos permitem explorar o agrupamento espacial em toda a área de LA.

Primeiro temos que atribuir neighbours a cada um dos distritos. Existem 2 métodos para fazer isso: Rook ou Queen. Podemos traçar as ligações entre os distritos vizinhos para visualizar sua distribuição no espaço.

```
neighbours <- poly2nb(districts)
neighbours

## Neighbour list object:
## Number of regions: 1135
## Number of nonzero links: 7080
## Percentage nonzero weights: 0.5495934
## Average number of links: 6.237885

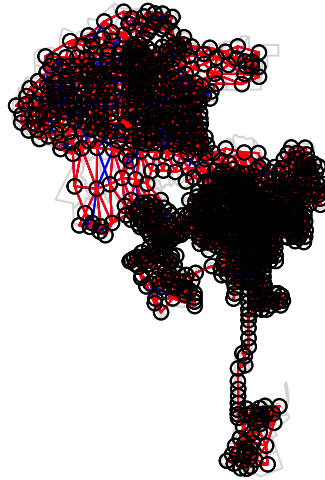
plot(districts, border = 'lightgrey')
plot(neighbours, coordinates(districts), add=TRUE, col='red')
```



```
neighbours2 <- poly2nb(districts, queen = FALSE)
neighbours2

## Neighbour list object:
## Number of regions: 1135
## Number of nonzero links: 5394
## Percentage nonzero weights: 0.4187157
## Average number of links: 4.752423

plot(districts, border = 'lightgrey')
plot(neighbours, coordinates(districts), add=TRUE, col='blue')
plot(neighbours2, coordinates(districts), add=TRUE, col='red')
```



Podemos ver que o segundo método (Rook) retorna muito menos links e podemos interpretar visualmente a diferença.

Autocorrelação Espacial Global

Agora que temos vizinhos, precisamos convertê-los.

Então podemos executar o teste de um moran. Isso cria um valor de correlação entre -1 e 1, a distribuição espacial desses valores é vista na figura da Introdução. 1 representa autocorrelação espacial positiva perfeita, 0 mostra que os dados são distribuídos aleatoriamente e -1 é autocorrelação espacial negativa.

```
listw <- nb2listw(neighbours)
listw

## Characteristics of weights list object:
## Neighbour list object:
## Number of regions: 1135
## Number of nonzero links: 7080
## Percentage nonzero weights: 0.5495934
## Average number of links: 6.237885
##
## Weights style: W
## Weights constants summary:
##      n      nn  S0      S1      S2
## W 1135 1288225 1135 392.1103 4642.774

moran.test(districts$Freq, listw)

##
```

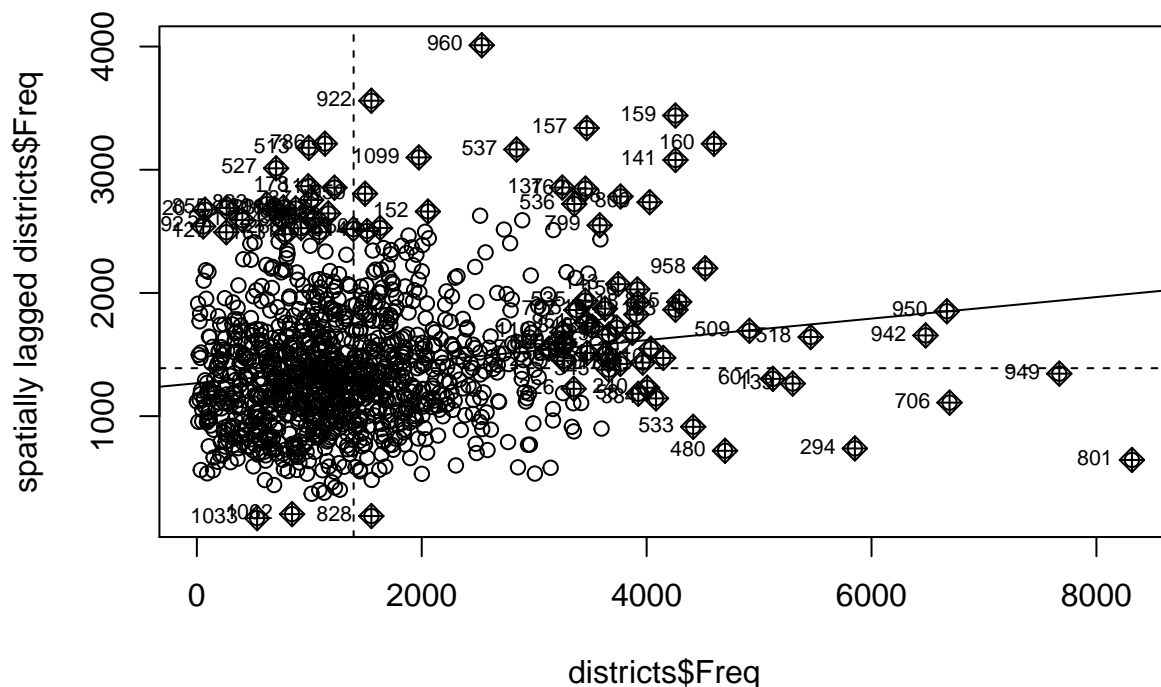
```
## Moran I test under randomisation
##
## data: districts$Freq
## weights: listw
##
## Moran I statistic standard deviate = 4.97, p-value = 3.348e-07
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##      0.0853613741      -0.0008818342      0.0003011199
```

A estatística Global Moran I é de 0,085, mostrando que a frequência de crimes em cada distrito é distribuída de forma significativamente aleatória. No entanto, vamos dar uma olhada nisso e ver se a autocorrelação espacial local retorna um resultado similar.

Autocorrelação Espacial Local

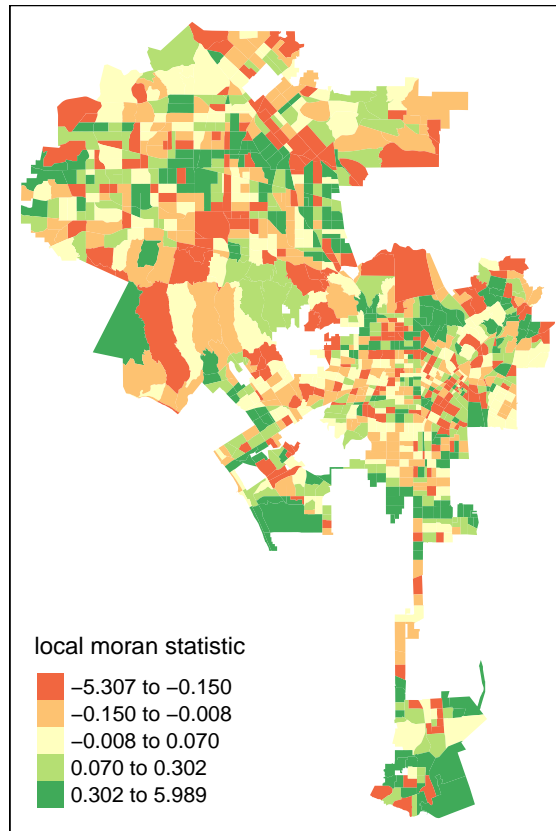
Em primeiro lugar, criamos um gráfico moran olhando para cada um dos valores plotados em relação aos seus valores espacialmente defasados. Explora a relação entre os dados e seus vizinhos.

```
moran <- moran.plot(districts$Freq, listw = nb2listw(neighbours2, style = "W"))
```



A partir do gráfico de dispersão, sugere potencialmente que a frequência dos crimes de cada distrito seja aleatória. Vamos tentar mapeá-lo.

```
local <- localmoran(x = districts$Freq, listw = nb2listw(neighbours2, style = "W"))
moran.map <- cbind(districts, local)
tm_shape(moran.map) + tm_fill(col = "li", style = "quantile", title = "local moran statistic")
```

Um valor positivo indica que o distrito é cercado por distritos de valor semelhante.

Do mapa, podemos ver que, em grande parte, nos distritos de LA, não há relação entre o número de crimes de um distrito e o de um distrito vizinho. No entanto, podemos ver alguns grupos de alta similaridade nos distritos do norte. Embora, não seja possível interpretar se estes são grupos de alta contagem de crimes ou baixos, apenas que são semelhantes. Então, nós conduzimos um mapa de cluster LISA.

LISA

```
quadrant <- vector(mode="numeric",length=nrow(local))

# Centraliza a variável de interesse em torno de sua média.
m.crime <- districts$Freq - mean(districts$Freq)

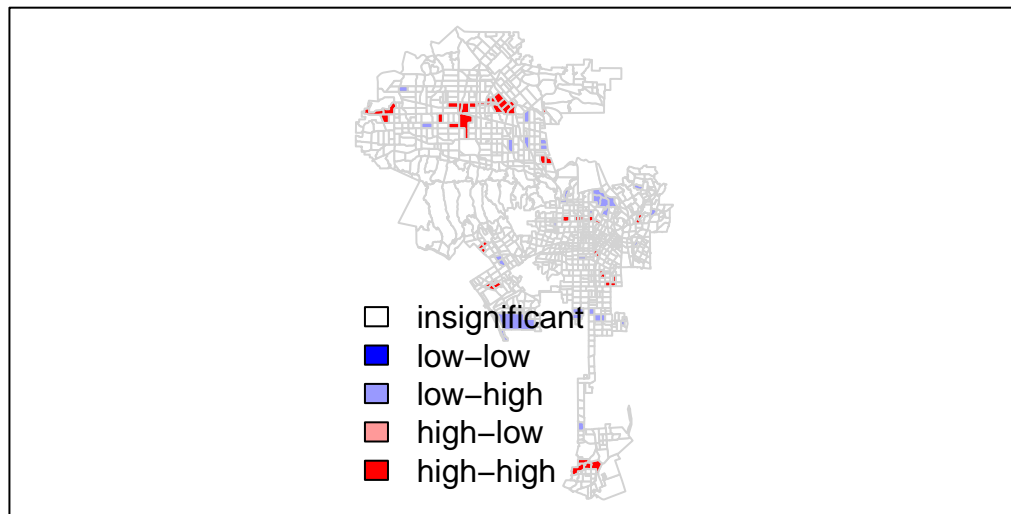
# Centraliza Moran's local em torno da média.
m.local <- local[,1] - mean(local[,1])

# Limiar de significância.
signif <- 0.1

# Constrói um quadrante de dados
quadrant[m.crime >0 & m.local>0] <- 4
quadrant[m.crime <0 & m.local<0] <- 1
quadrant[m.crime <0 & m.local>0] <- 2
quadrant[m.crime >0 & m.local<0] <- 3
quadrant[local[,5]>signif] <- 0
```

```
brks <- c(0,1,2,3,4)
colors <- c("white","blue",rgb(0,0,1,alpha=0.4),rgb(1,0,0,alpha=0.4),"red")

plot(districts,border="lightgray",col=colors[findInterval(quadrant,brks,all.inside=FALSE)])
box()
legend(x = -118.7, y = 34,legend=c("insignificant","low-low","low-high","high-low","high-high"),
      fill=colors,bty="n")
```



Conclusão

Como podemos ver no mapa, que o aglomerado, identificado no mapa anterior do moran, eram distritos de altas contagens de crimes relatados cercados por outros distritos de altos relatórios de crimes. Surpreendentemente, 2 das variáveis do cluster não estão presentes: distritos de baixa frequência do crime cercados por distritos similares de baixa incidência de crimes (azul escuro) e distritos de alta criminalidade cercados por distritos de baixa frequência criminal.

Disponibilidade dos dados

O conjunto de dados bem como o código R e arquivos auxiliares utilizados nesta análise podem ser encontrados em: <https://github.com/erivandoramos/Crime-em-Los-Angeles>

Referências

Ambarish Ganguly. **Little Book on Data Visualization and Modelling**. Disponível em: <<https://ambarishg.github.io/books/LittleBookDataViz>> Acesso em: 22 mai. 2019.