Universidad de los Andes

Modelos Estadísticos de Crimen

27 de Mayo de 2015

Motivacion

Conocer la densidad de probabilidad del crimen sobre una ciudad, a distintas horas del día, permite

Motivacion

Conocer la densidad de probabilidad del crimen sobre una ciudad, a distintas horas del día, permite

- Establecer el efecto de la presencia de estaciones de policía y patrullajes.

Motivacion

Conocer la densidad de probabilidad del crimen sobre una ciudad, a distintas horas del día, permite

- Establecer el efecto de la presencia de estaciones de policía y patrullajes.
- Determinar factores de la geografia que incrementan o disminuyen la intensidad de crimen.

Motivacion

Conocer la densidad de probabilidad del crimen sobre una ciudad, a distintas horas del día, permite

- Establecer el efecto de la presencia de estaciones de policía y patrullajes.
- Determinar factores de la geografia que incrementan o disminuyen la intensidad de crimen.
- Determinar los hotspots y su movimiento a lo largo del día

Motivacion

Conocer la densidad de probabilidad del crimen sobre una ciudad, a distintas horas del día, permite

- Establecer el efecto de la presencia de estaciones de policía y patrullajes.
- Determinar factores de la geografia que incrementan o disminuyen la intensidad de crimen.
- Determinar los hotspots y su movimiento a lo largo del día
- Diseñar patrullajes óptimos

Motivacion

Estimar la densidad del crimen a partir de los datos es un reto por

- La alta dimensionalidad de los datos
- El tamaño de la base

Procesos Puntuales de Poisson

- Un proceso puntual de Poisson, con función de intensidad μ, es un proceso por el que un conjunto de puntos son generados aleatoriamente en $\mathbb{R}^n$.
- Para un conjunto $S \subseteq \mathbb{R}^n$ defina N_S como el número de puntos que caen en S. Entonces, un proceso de Poisson debe satisfacer
  1. Si $A \cap B = \emptyset$ entonces N_A y N_B son independientes.
  2. El número de puntos que cae en S está dado por una distribución de Poisson con intensidad $\mu(A)$

$$Pr(N_A = n) = \frac{\mu(A)^n e^{\mu(A)}}{n!} \tag{1}$$

Procesos Puntuales de Poisson

Los PPP son usados para simular fenómenos como

Procesos Puntuales de Poisson

Los PPP son usados para simular fenómenos como

- Llegada de autobuses

Procesos Puntuales de Poisson

Los PPP son usados para simular fenómenos como

- Llegada de autobuses
- Ocurrencia de terremotos

Procesos Puntuales de Poisson

Los PPP son usados para simular fenómenos como

- Llegada de autobuses
- Ocurrencia de terremotos
- Ubicación de asentamientos humanos

Procesos Puntuales de Poisson

Los PPP son usados para simular fenómenos como

- Llegada de autobuses
- Ocurrencia de terremotos
- Ubicación de asentamientos humanos
- Posición de las estrellas

Procesos Puntuales de Poisson

Sus ventajas radican es que es un buen equilibrio entre complejidad y facilidad de estimación, pero no es completamente adecuado para simular ocurrencia de crímenes.

- La intensidad del crimen cambia y se desplaza a lo largo del tiempo
- Los crímenes no son independientes, sino que crimen llama más crimen (”efecto de las ventanas rotas”)

Estimación de un PPP

- Dada una realización de un PPP, como la ocurrencia de hurtos en Bogotá durante el 2013, es posible estimar la intensidad del proceso.

Estimación de un PPP

- Dada una realización de un PPP, como la ocurrencia de hurtos en Bogotá durante el 2013, es posible estimar la intensidad del proceso.
- Este proceso vive en $\mathbb{R}^3$, donde las dimensiones son: Latitud, Longitud y Tiempo.

Estimación de un PPP

- Dada una realización de un PPP, como la ocurrencia de hurtos en Bogotá durante el 2013, es posible estimar la intensidad del proceso.
- Este proceso vive en $\mathbb{R}^3$, donde las dimensiones son: Latitud, Longitud y Tiempo.
- No se encontró un paquete en R capaz de hacer la estimación en más de una dimensión y que usara algún algoritmo de aprendizaje.

Estimación de un PPP

- Dada una realización de un PPP, como la ocurrencia de hurtos en Bogotá durante el 2013, es posible estimar la intensidad del proceso.
- Este proceso vive en $\mathbb{R}^3$, donde las dimensiones son: Latitud, Longitud y Tiempo.
- No se encontró un paquete en R capaz de hacer la estimación en más de una dimensión y que usara algún algoritmo de aprendizaje.
- Se aproximará la intensidad usando estimación de densidades por Kernel, donde los parámetros óptimos del ancho de banda se encontrarán usando máxima verosimilitud.

Estimación por Kernels

- Suponemos que la intensidad μ es una medida con una densidad f, de tal forma que

$$\mu(A) = \int_A f(x)dx \tag{2}$$

Estimación por Kernels

- Suponemos que la intensidad μ es una medida con una densidad f, de tal forma que

$$\mu(A) = \int_A f(x)dx \tag{2}$$

- Dada una muestra de Test $Y = \{y_k\}$, la verosimilitud de f estará dada por

$$L(f) = \prod_{k=1}^{Te} f(y_k) \tag{3}$$

Estimación por Kernels

- Es necesario definir una familia de funciones de la cual escoger f.

Estimación por Kernels

- Es necesario definir una familia de funciones de la cual escoger f.
- Asumiremos que f es un kernel, obtenido a partir de observaciones de una base de entrenamiento $\{x_l\}$. Esto quiere decir que

$$f(y) = \frac{1}{Tr} \sum_{l=1}^{Tr} k(y; x_l, \Sigma) \tag{4}$$

donde $k(\cdot\,; x, \Sigma)$ es un kernel centrado en x con ancho de banda $\Sigma \in \mathbb{R}^{d \times d}$.

Estimación por Kernels

La función de log-verosimilitud estará dada por

$$LL(\Sigma) = \sum_{k=1}^{Te} log(f(y_k)) \quad (5)$$

$$= \frac{1}{Te} \sum_{k=1}^{Te} log\left(\frac{1}{Tr} \sum_{l=1}^{Tr} k(y_k; x_l, \Sigma) \right) \quad (6)$$

Estimación por Kernels

Debemos encontrar el ancho de banda que maximice LL. Para esto usamos el método de descenso del gradiente. Note que si θ es alguno de los parámetros de los que depende LL, entonces

$$\frac{\partial LL}{\partial \theta} = \frac{1}{Te} \sum_{k=1}^{Te} \frac{\sum_{l=1}^{Tr} \frac{\partial k(y_k; x_l, \Sigma)}{\partial \theta}}{\sum_{l=1}^{Tr} k(y_k; x_l, \Sigma)} \tag{7}$$

Estimación por Kernels

- Escogemos un kernel gaussiano, pues permite calcular las derivadas de LL de forma sencilla.

$$k(y;x,\Sigma) = \frac{1}{(\sqrt{2\pi})^d\sqrt{|\Sigma|}} e^{-\frac{1}{2}(y-x)^T\Sigma(y-x)} \tag{8}$$

Estimación por Kernels

- Escogemos un kernel gaussiano, pues permite calcular las derivadas de LL de forma sencilla.

$$k(y;x,\Sigma) = \frac{1}{(\sqrt{2\pi})^d\sqrt{|\Sigma|}} e^{-\frac{1}{2}(y-x)^T\Sigma(y-x)} \tag{8}$$

- Asumiremos que $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_d)$, de forma que

$$k(y;x,\sigma) = \frac{1}{(\sqrt{2\pi})^d\sigma_1\ldots\sigma_d} e^{-\frac{1}{2}\sum_{i=1}^{d}\left[\frac{y_i-x_i}{\sigma_i}\right]^2} \tag{9}$$

Estimación por Kernels

- Usando la densidad gaussiana k y la fórmula para $\frac{\partial LL}{\partial \theta}$ obtenemos

$$\frac{\partial LL}{\partial \sigma_j} = \frac{1}{Te}\frac{1}{\sigma_j}\sum_{k=1}^{Te}\frac{\sum_{l=1}^{Tr} k(y_k; x_l, \sigma)\left[\left(\frac{y_k^{(j)} - x_l^{(j)}}{\sigma_j^2}\right) - 1\right]}{\sum_{l=1}^{Tr} k(y_k; x_l, \sigma)} \tag{10}$$

Estimación por Kernels

- Usando la densidad gaussiana k y la fórmula para $\frac{\partial LL}{\partial \theta}$ obtenemos

$$\frac{\partial LL}{\partial \sigma_j} = \frac{1}{Te}\frac{1}{\sigma_j}\sum_{k=1}^{Te} \frac{\sum_{l=1}^{Tr} k(y_k; x_l, \sigma)\left[\left(\frac{y_k^{(j)} - x_l^{(j)}}{\sigma_j^2}\right) - 1\right]}{\sum_{l=1}^{Tr} k(y_k; x_l, \sigma)} \tag{10}$$

- Y el gradiente de la función de log-verosimilitud estará dado por

$$\nabla LL = \left(\frac{\partial LL}{\partial \sigma_1}, \ldots, \frac{\partial LL}{\partial \sigma_d}\right) \tag{11}$$

Descenso del gradiente

- El método de descenso del gradiente para maximizar una función $f(x)$ consiste en iniciar con un guess del x óptimo, y actualizarlo moviéndose en la dirección ∇f.

Descenso del gradiente

- El método de descenso del gradiente para maximizar una función $f(x)$ consiste en iniciar con un guess del x óptimo, y actualizarlo moviéndose en la dirección ∇f.
- En el caso de $LL(\sigma)$ iniciamos con un guess σ_1 del ancho de banda óptimo, y lo actualizamos según la regla

$$\sigma_{k+1} = \sigma_k + \eta_k \nabla LL(\sigma_k) \tag{12}$$

Descenso del gradiente

- El método de descenso del gradiente para maximizar una función $f(x)$ consiste en iniciar con un guess del x óptimo, y actualizarlo moviéndose en la dirección ∇f.
- En el caso de $LL(\sigma)$ iniciamos con un guess σ_1 del ancho de banda óptimo, y lo actualizamos según la regla

$$\sigma_{k+1} = \sigma_k + \eta_k \nabla LL(\sigma_k) \tag{12}$$

- η_k es llamada la tasa de aprendizaje. Para garantizar convergencia a un óptimo local basta que

$$\lim_{k\to\infty} \eta_k = 0 \qquad \sum_{k=1}^{\infty} \eta_k = \infty \tag{13}$$

Descenso del gradiente

- Para escoger la tasa de aprendizaje usamos la técnica de Bold Driver.

Descenso del gradiente

- Para escoger la tasa de aprendizaje usamos la técnica de Bold Driver.
- Si en la iteración $\sigma_{k+1} = \sigma_k + \eta_k \nabla LL(\sigma_k)$ la función LL aumenta, entonces la tasa de aprendizaje η_k será aumentada en 5%.

Descenso del gradiente

- Para escoger la tasa de aprendizaje usamos la técnica de Bold Driver.
- Si en la iteración $\sigma_{k+1} = \sigma_k + \eta_k \nabla LL(\sigma_k)$ la función LL aumenta, entonces la tasa de aprendizaje η_k será aumentada en 5%.
- Si por el contrario LL disminuyó, no actualizamos σ_k, sino que disminuimos la tasa de aprendizaje en un 50%

Descenso del gradiente

- Para escoger la tasa de aprendizaje usamos la técnica de Bold Driver.
- Si en la iteración $\sigma_{k+1} = \sigma_k + \eta_k \nabla LL(\sigma_k)$ la función LL aumenta, entonces la tasa de aprendizaje η_k será aumentada en 5%.
- Si por el contrario LL disminuyó, no actualizamos σ_k, sino que disminuimos la tasa de aprendizaje en un 50%
- Cuál es el guess inicial?

Descenso del gradiente

- Para escoger la tasa de aprendizaje usamos la técnica de Bold Driver.
- Si en la iteración $\sigma_{k+1} = \sigma_k + \eta_k \nabla LL(\sigma_k)$ la función LL aumenta, entonces la tasa de aprendizaje η_k será aumentada en 5%.
- Si por el contrario LL disminuyó, no actualizamos σ_k, sino que disminuimos la tasa de aprendizaje en un 50%
- Cuál es el guess inicial?
- Cuándo parar?

Descenso del gradiente

- Para escoger la tasa de aprendizaje usamos la técnica de Bold Driver.
- Si en la iteración $\sigma_{k+1} = \sigma_k + \eta_k \nabla LL(\sigma_k)$ la función LL aumenta, entonces la tasa de aprendizaje η_k será aumentada en 5%.
- Si por el contrario LL disminuyó, no actualizamos σ_k, sino que disminuimos la tasa de aprendizaje en un 50%
- Cuál es el guess inicial?
- Cuándo parar?
- El problema será cóncavo?

Base de Datos

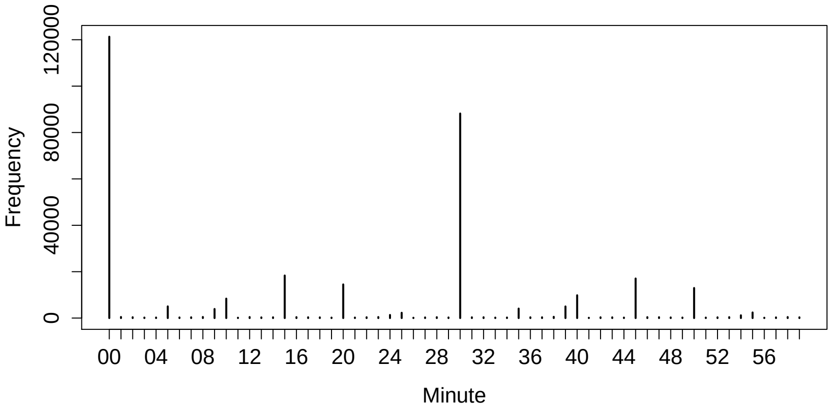
Delito	Frecuencia	Participación	Tipo
1	9977	3.0%	Homicidio Doloso
2	3526	1.1%	Homicidio Culposo
3	60539	18.4%	Lesiones Dolosas
4	18982	5.8%	Lesiones Culposas
5	39570	12.0%	Hurto Residencias
6	11519	3.5%	Hurto Motos
7	24396	7.4%	Hurto Carros
9	115794	35.1%	Hurto General
14	45490	13.8%	Estupefacientes

Base de Datos

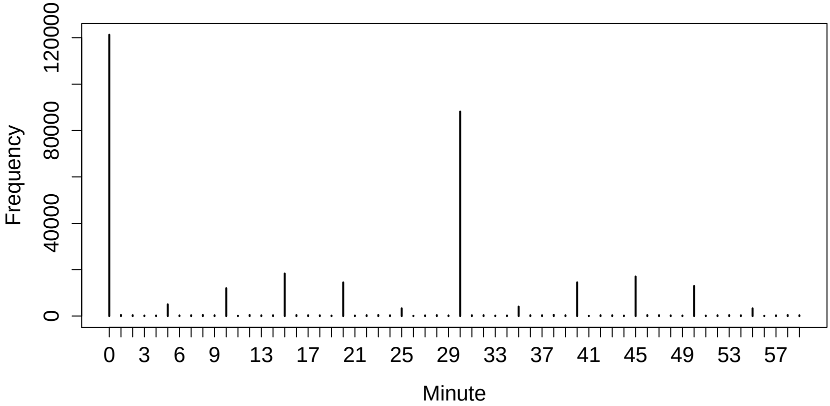
Decisiones que hay que tomar para estimar una densidad:

- Cómo corregir las horas de los delitos?
- Qué intervalo de tiempo usar?
- Agrupar los crímenes?

Base de Datos



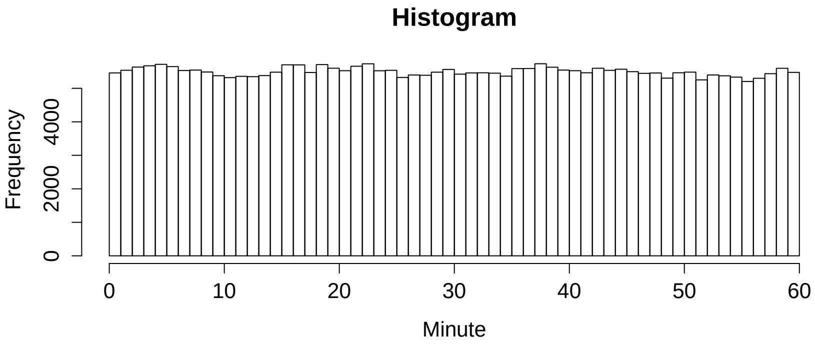
Base de Datos



Base de Datos

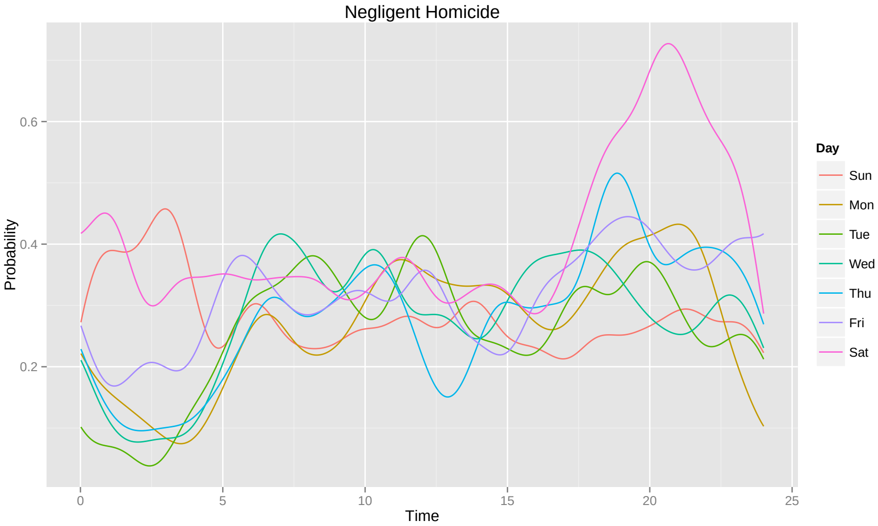
```
Call:
lm(formula = objeto ~ ., data = Model)

Residuals:
     Min       1Q   Median       3Q      Max 
-1504.25   -82.83     0.42    61.92  1101.25 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   315.58      55.09   5.728 4.63e-07 ***
hour        33098.00     539.78  61.317  < 2e-16 ***
half        60937.00     539.78 112.891  < 2e-16 ***
fifteen     13762.75     330.55  41.636  < 2e-16 ***
ten          9559.50     269.89  35.420  < 2e-16 ***
five         3623.17     198.64  18.240  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 381.7 on 54 degrees of freedom
Multiple R-squared:  0.9996, Adjusted R-squared:  0.9996 
F-statistic: 3.034e+04 on 5 and 54 DF,  p-value: < 2.2e-16
```

Base de Datos



Densidad de los Delitos en el Tiempo

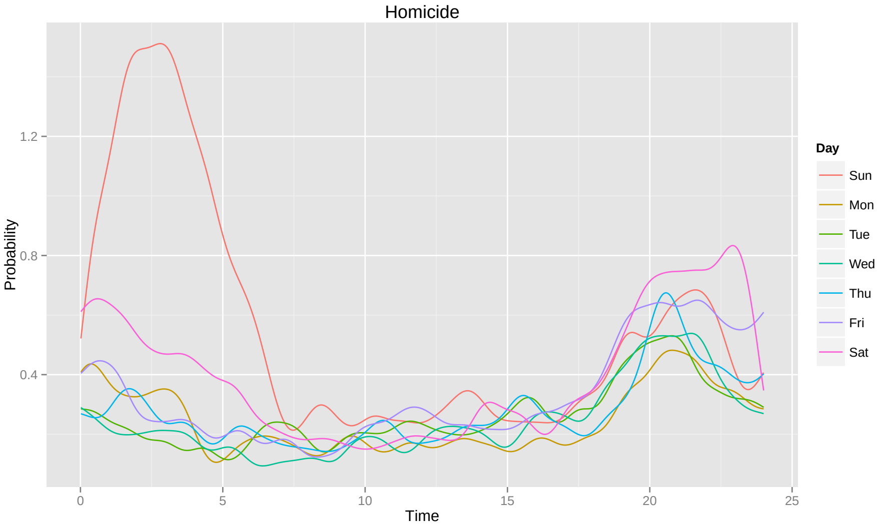
- Qué día y hora es más probable que ocurran homicidios accidentales?



Negligent Homicide
Probability
0.6
0.4
0.2
0
5
10
15
20
25
Time
Day
Sun
Mon
Tue
Wed
Thu
Fri
Sat

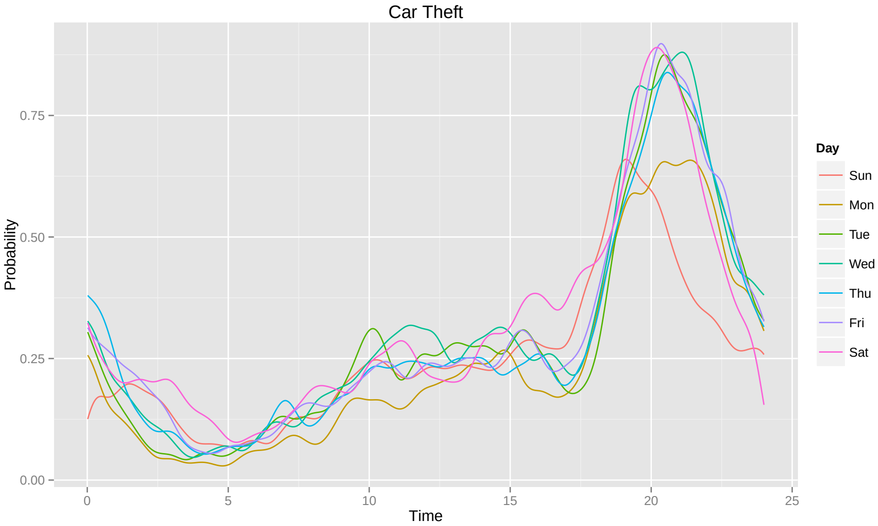
Densidad de los Delitos en el Tiempo

- Qué día y hora es más probable que ocurran homicidios accidentales?
- Homicidios no accidentales?


Homicide
Probability
1.2
0.8
0.4
0
5
10
15
20
25
Time
Day
Sun
Mon
Tue
Wed
Thu
Fri
Sat

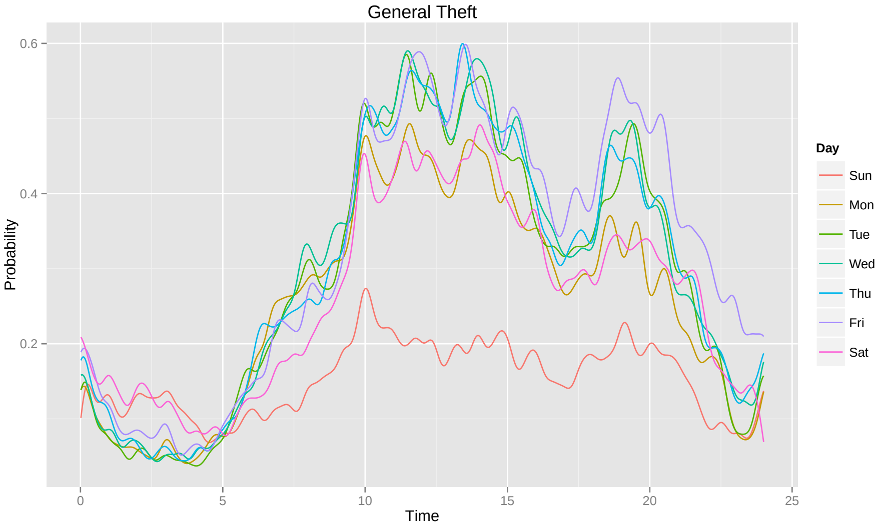
Densidad de los Delitos en el Tiempo

- Qué día y hora es más probable que ocurran homicidios accidentales?
- Homicidios no accidentales?
- Qué día es menos probable que se le roben el carro?


Car Theft
Probability
0.75
0.50
0.25
0.00
0
5
10
15
20
25
Time
Day
Sun
Mon
Tue
Wed
Thu
Fri
Sat

Densidad de los Delitos en el Tiempo

- Qué día y hora es más probable que ocurran homicidios accidentales?
- Homicidios no accidentales?
- Qué día es menos probable que se le roben el carro?
- Cuál es el horario de trabajo de los atracadores?


General Theft
Probability
Time
0.6
0.4
0.2
0
5
10
15
20
25
Day
Sun
Mon
Tue
Wed
Thu
Fri
Sat

Densidad de los Delitos en el Espacio

- Correr el descenso del gradiente en el espacio y el tiempo toma mucho más tiempo. 2h-1dia

Densidad de los Delitos en el Espacio

- Correr el descenso del gradiente en el espacio y el tiempo toma mucho más tiempo. 2h-1dia
- A esta hora, cuál es el punto más probable en el que ocurrirá un homicidio accidental?

Densidad de los Delitos en el Espacio

- Correr el descenso del gradiente en el espacio y el tiempo toma mucho más tiempo. 2h-1dia
- A esta hora, cuál es el punto más probable en el que ocurrirá un homicidio accidental? (4.684444, -74.096000)

Densidad de los Delitos en el Espacio

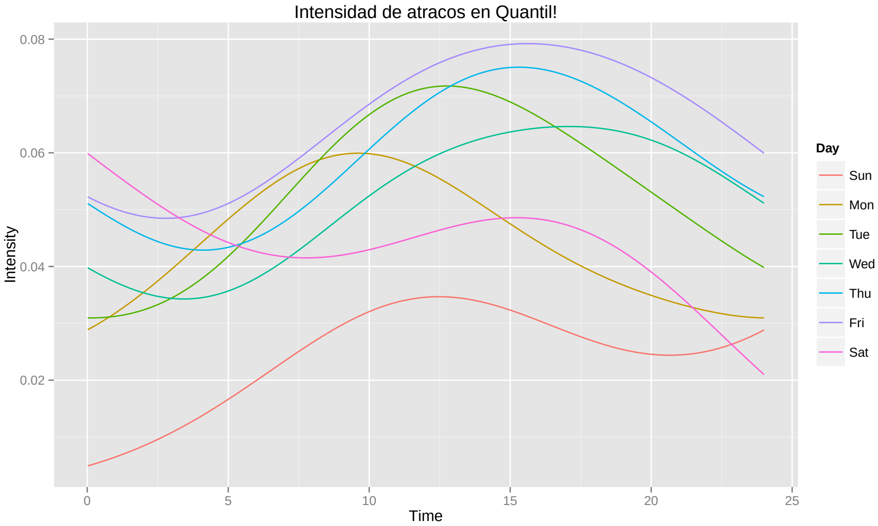
- Correr el descenso del gradiente en el espacio y el tiempo toma mucho más tiempo. 2h-1dia
- A esta hora, cuál es el punto más probable en el que ocurrirá un homicidio accidental? (4.684444, -74.096000)
- Cuál es el punto más probable donde ocurrirá un homicidio?

Densidad de los Delitos en el Espacio

- Correr el descenso del gradiente en el espacio y el tiempo toma mucho más tiempo. 2h-1dia
- A esta hora, cuál es el punto más probable en el que ocurrirá un homicidio accidental? (4.684444, -74.096000)
- Cuál es el punto más probable donde ocurrirá un homicidio? (4.607486,-74.078245)

Densidad de los Delitos en el Espacio

- Correr el descenso del gradiente en el espacio y el tiempo toma mucho más tiempo. 2h-1dia
- A esta hora, cuál es el punto más probable en el que ocurrirá un homicidio accidental? (4.684444, -74.096000)
- Cuál es el punto más probable donde ocurrirá un homicidio? (4.607486,-74.078245)
- A qué hora salir de Quantil para que no lo atraquen?



Intensidad de atracos en Quantil!
Intensity
0.08
0.06
0.04
0.02
Time
0
5
10
15
20
25
Day
Sun
Mon
Tue
Wed
Thu
Fri
Sat

Próximos Desarrollos

- Estudiar la relación entre elementos de la geografía y la intensidad del crimen.

Próximos Desarrollos

- Estudiar la relación entre elementos de la geografía y la intensidad del crimen.
- Escribir el código en otro lenguaje más eficiente: Matlab, Python?

Próximos Desarrollos

- Estudiar la relación entre elementos de la geografía y la intensidad del crimen.
- Escribir el código en otro lenguaje más eficiente: Matlab, Python?
- Encontrar una plataforma más amigable para visualizar los datos de forma dinámica: ArcGis, Java?

Próximos Desarrollos

- Estudiar la relación entre elementos de la geografía y la intensidad del crimen.
- Escribir el código en otro lenguaje más eficiente: Matlab, Python?
- Encontrar una plataforma más amigable para visualizar los datos de forma dinámica: ArcGis, Java?
- Posibles aplicaciones: Encontrar la ruta más segura entre dos puntos de la ciudad, diseñar un patrullaje óptimo.