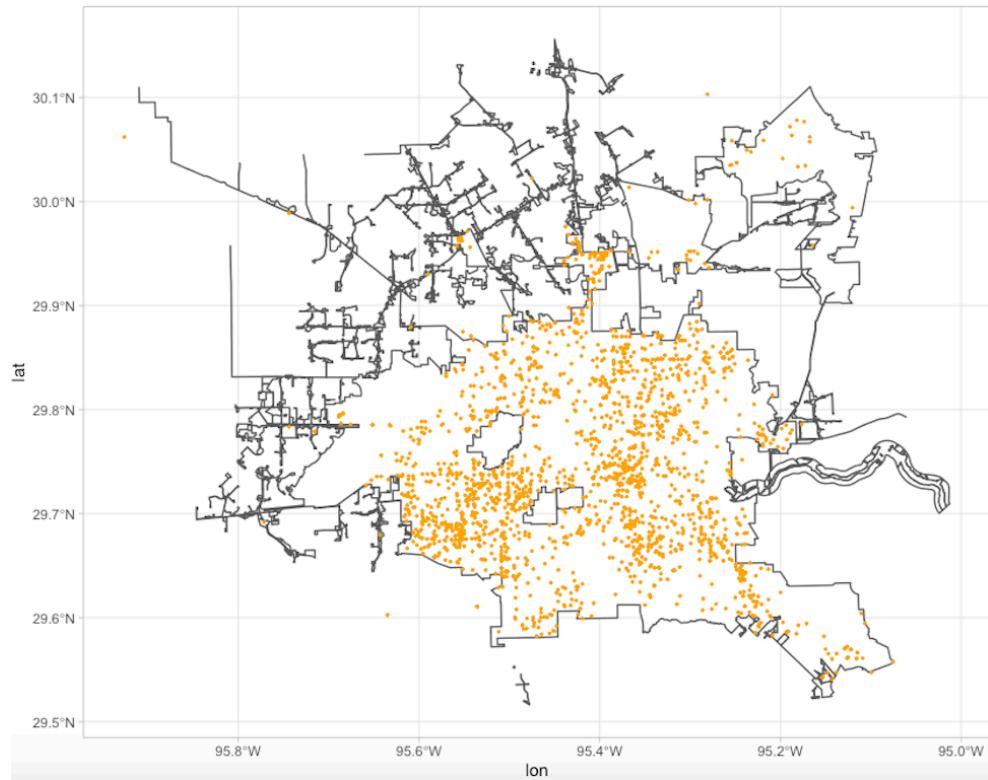


MAPEAMENTO E ANÁLISE DE CRIMES OCORRIDOS EM HOUSTON

NO PERÍODO DE JANEIRO A AGOSTO DE 2010



MBA Executivo em Business Analytics e Big Data
T8 – Estatística Espacial – Prof. Jony Arrais

Rafael Rocha
A56660250

MAPEAMENTO E ANÁLISE DE CRIMES OCORRIDOS EM HOUSTON

JAN A AGO DE 2010

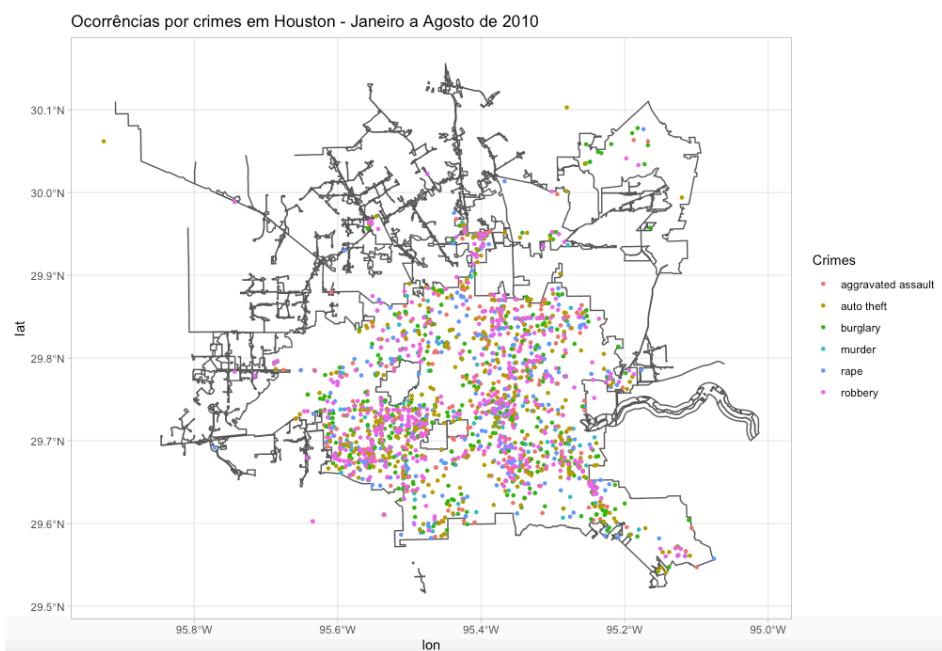
OBJETIVO DO PROJETO

Com base nos dados, com localização exata de latitude e longitude, de crimes diversos ocorridos na populosa cidade de Houston, do estado de Texas dos Estados Unidos, tais como: ***aggravated assault, auto theft, burglary, murder, rape e robbery***; este projeto tem como objetivo, fazer uma avaliação visual do comportamento de cada tipo de crime, tendo como janela os dois primeiros quadrimestres de 2010.

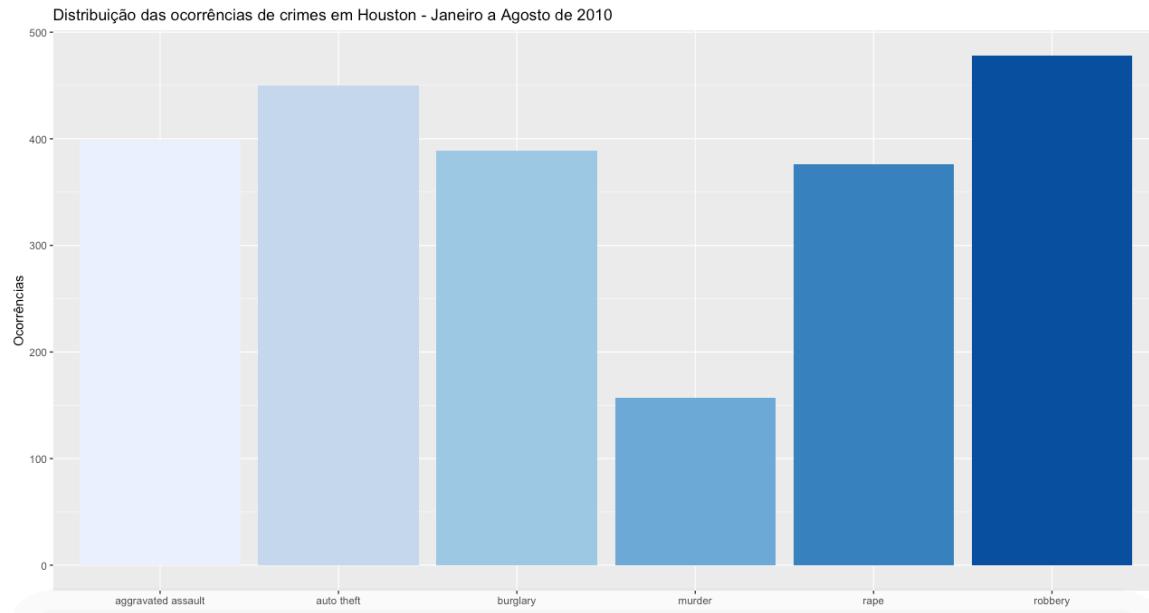
Serão apresentados mapas de pontos e investigados os efeitos de primeira e segunda ordem, para cada crime distinto, e serão discutidas diferenças observadas entre eles e os períodos escolhidos com uma análise exploratória e interpretação de resultados.

O objetivo desse relatório é dar suporte a autoridades locais na prevenção e redução da criminalidade.

OCORRÊNCIAS POR CRIMES



O foco foi nas colunas: **offense**, **lon** e **lat** da nossa base de dados, sendo essas de nosso maior interesse. Se nota que as coordenadas geográficas estão em latitude e longitude, que em resumo, formam uma grade sobre a Terra, nos possibilitando apontar a localização exata de qualquer local do globo.



A figura acima tem o intuito de facilitar a avaliação visual da distribuição das ocorrências de cada crime ao longo dos dois oito primeiros meses de 2010. Podemos verificar de não há uma diferença considerável entre os diversos crimes, *robbery* é o delito com mais casos e *murder* com o menor número.

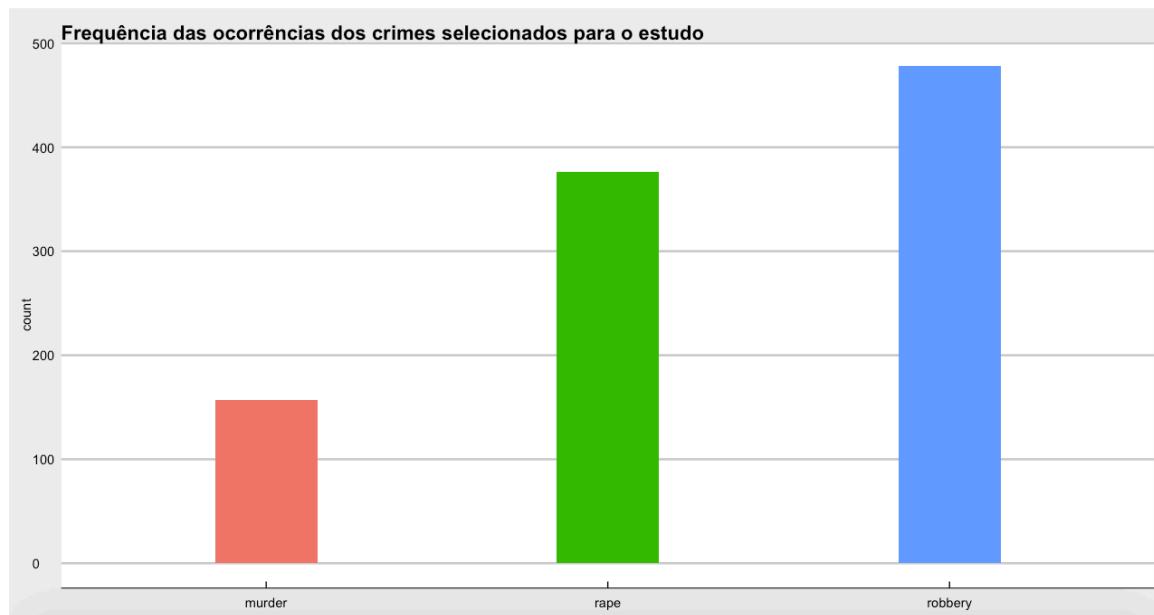


No gráfico acima, onde foi ilustrada a evolução do número de infrações, na janela dos oito primeiros meses de 2010, podemos notar que em maio atingimos o pico de ocorrências de delitos na cidade.

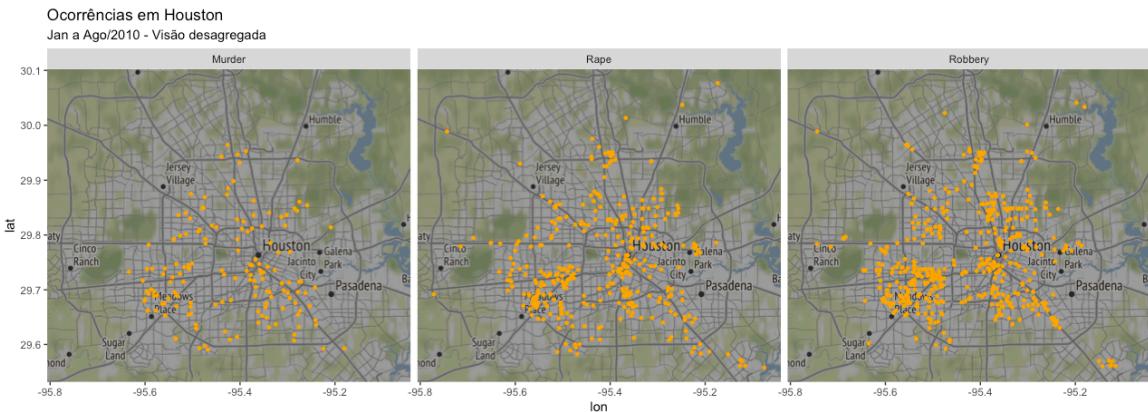
RAPE, ROBBERY E MURDER

Em seguida, vamos avaliar o comportamento de diferentes tipos de crimes, e para essa ocasião foram escolhidos *rape*, *robbery* e *murder*.

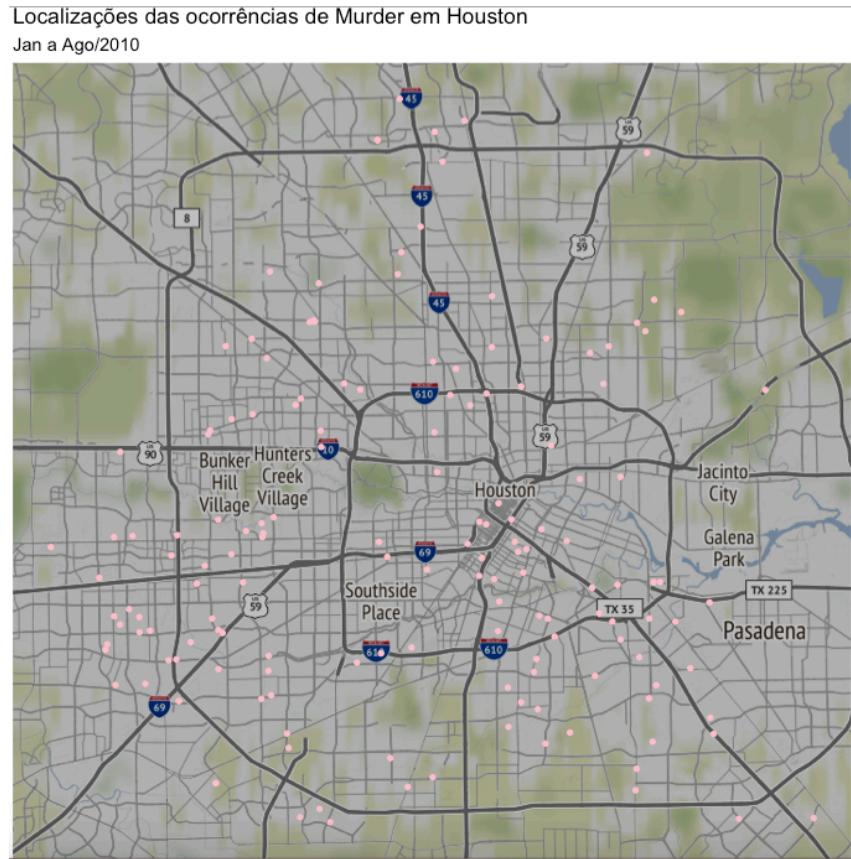
No próximo gráfico analisaremos a frequência desses três delitos independentes. *Murder* e *robbery* foram escolhidos por representarem os delitos com menor e maior número de infrações registrados, e *rape* por estar próximo a média se levarmos todos os seis em consideração.



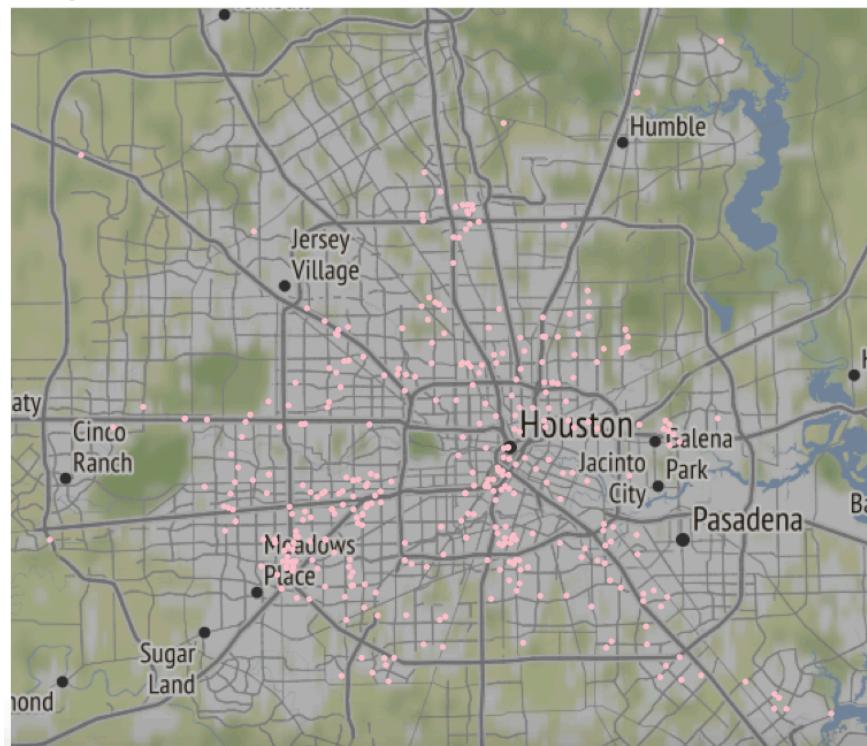
Em seguida serão apresentados mapas com padrões espaciais de pontos, compostos com as coordenadas dos eventos em questão. A análise espacial desses padrões nos possibilita examinar se um conjunto de eventos tem padrão sistemático ou aleatório.



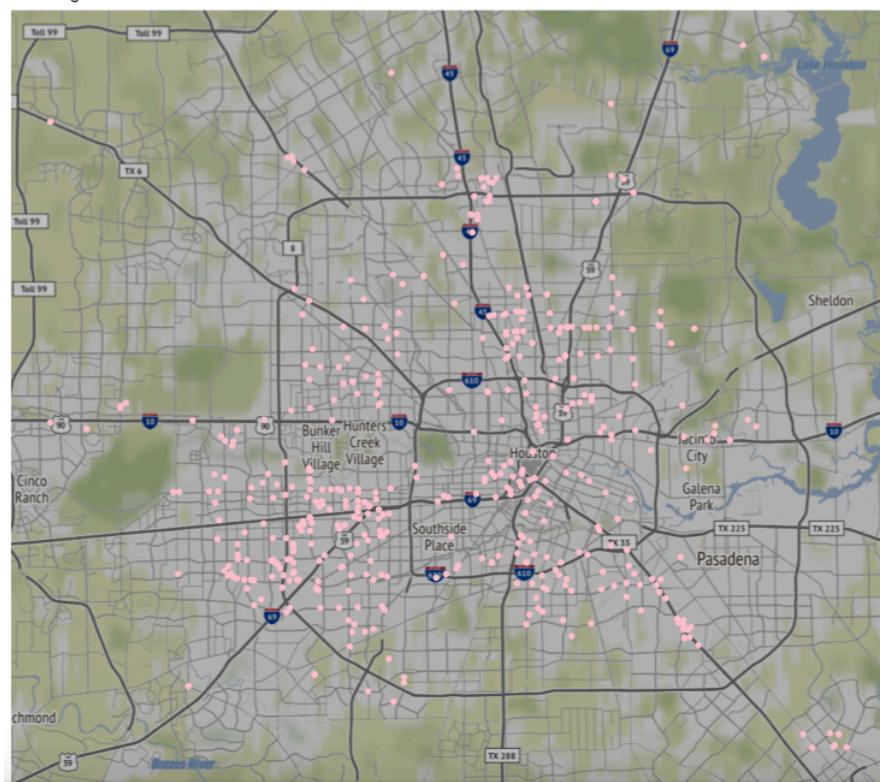
A quantidade de pontos amarelos confirma a frequência ilustrada no gráfico de barras apresentado anteriormente, onde *murder* tinha a menor e *robbery* o maior número de ocorrências. Serão ilustrados a seguir cada mapa separadamente para facilitar a visualização.



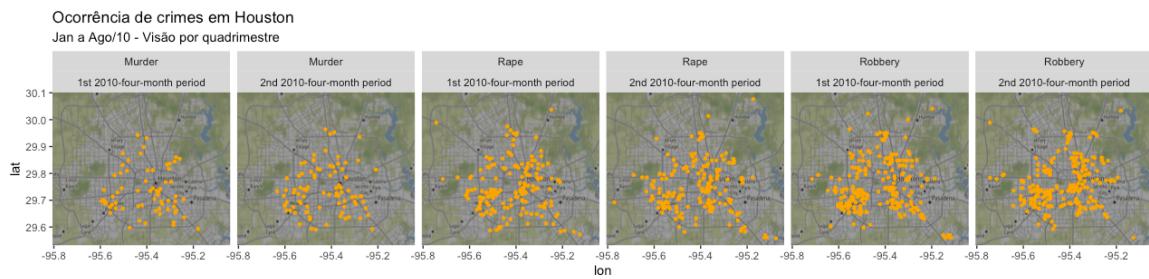
Localizações das ocorrências de Rape em Houston
Jan a Ago/2010



Localizações das ocorrências de Robbery em Houston
Jan a Ago/2010



A construção dos mapas acima nos faz pensar se existem padrões nos espaços, dentro de um cenário onde o problema é conhecido, no caso os crimes, mas ainda não se sabe se as localizações onde ocorreram esses fenômenos foram aleatórias. Vamos aprofundar esse estudo mais a frente tratando esses cenários para descobrir.



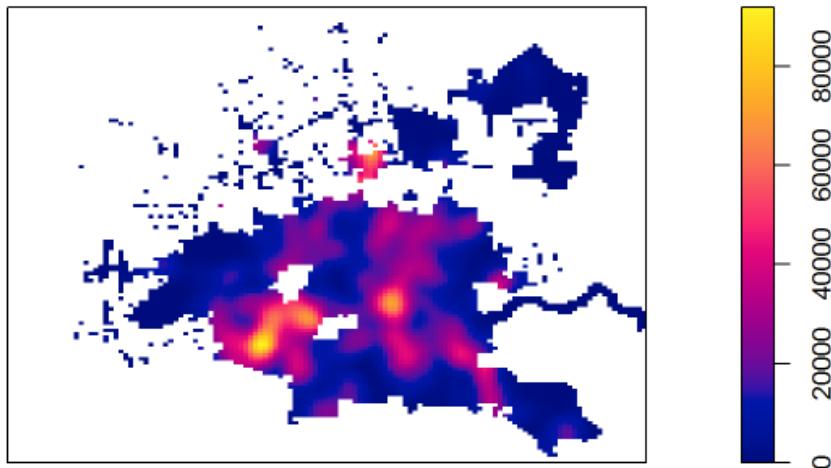
Nos mapas acima, os três tipos de crimes foram subdivididos em quadrimestres, de janeiro a abril e de maio a agosto. Mesmo que discretamente, podemos notar que todos os crimes tiveram aumento no segundo quadrimestre. As ocorrências de *murder* subiram mais de 30%, de *rape* pouco mais de 10% e de *robbery* quase 1%.

ESTIMAÇÃO DE EFEITOS DE PRIMEIRA ORDER

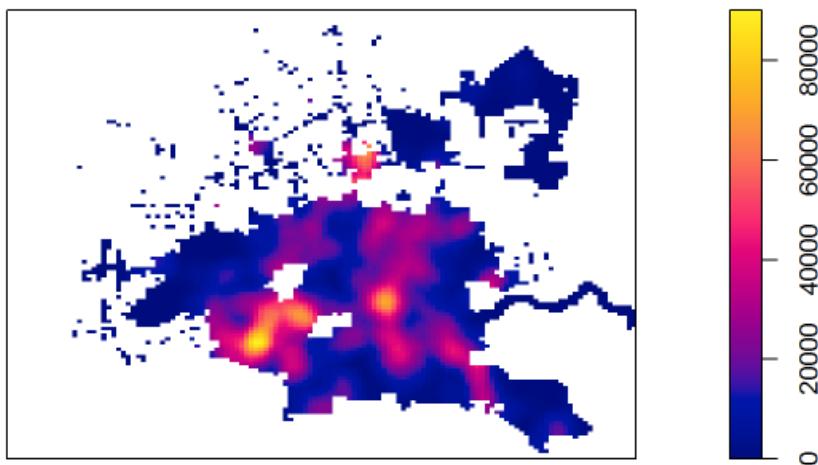
Vamos agora explorar os efeitos de primeira ordem e para dar suporte a nossa investigação, utilizaremos o estimador KERNEL, que é um método de interpolação que mede a intensidade das ocorrências na região de estudo.

Nos mapas abaixo, vamos verificar o comportamento espacial de todas as ocorrências de todos os seis delitos unificados, com foco na intensidade dos registros, perceba que os resultados dos dois diferentes estimadores de Kernel foram bastante parecidos. Algumas áreas de Houston, com as cores mais amareladas, independente do tipo de delito, apresentam maior aglomeração de incidentes, o que parece ser um padrão não importando necessariamente o tipo de crime. A intensidade não parece ser constante e neste caso podemos afirmar que não se trata de um processo estacionário.

Kernel Quartico

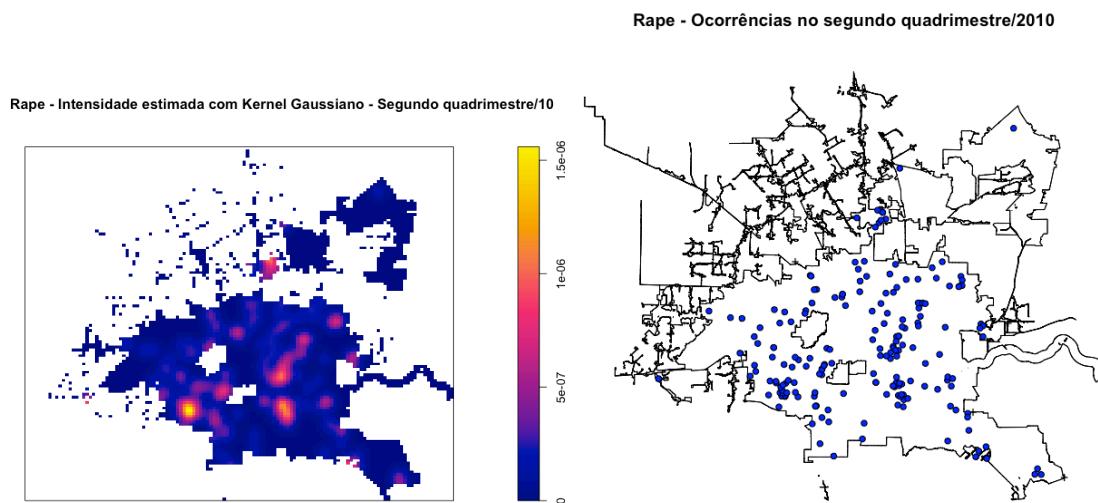
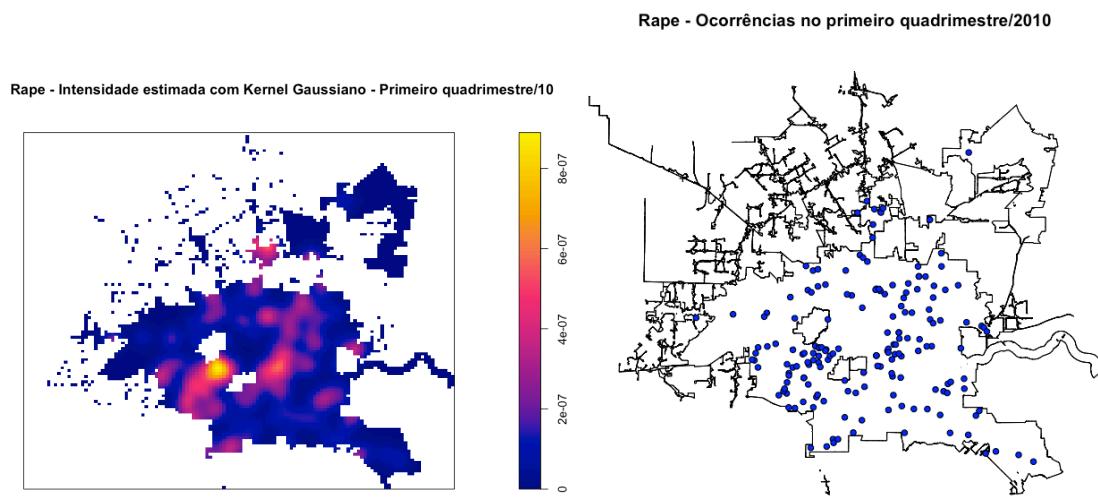


Kernel Epanechnikov

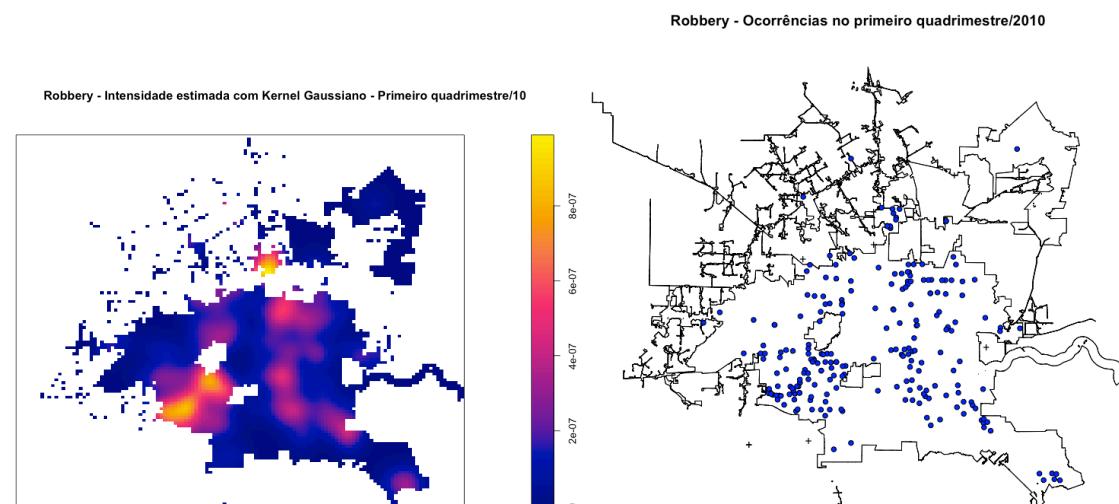


Verificaremos a partir desse momento, com aplicação do estimador de kernel gaussiano, os tipos de crimes divididos por quadrimestres e também um mapa de pontos para que possa ser feita a comparação dos resultados.

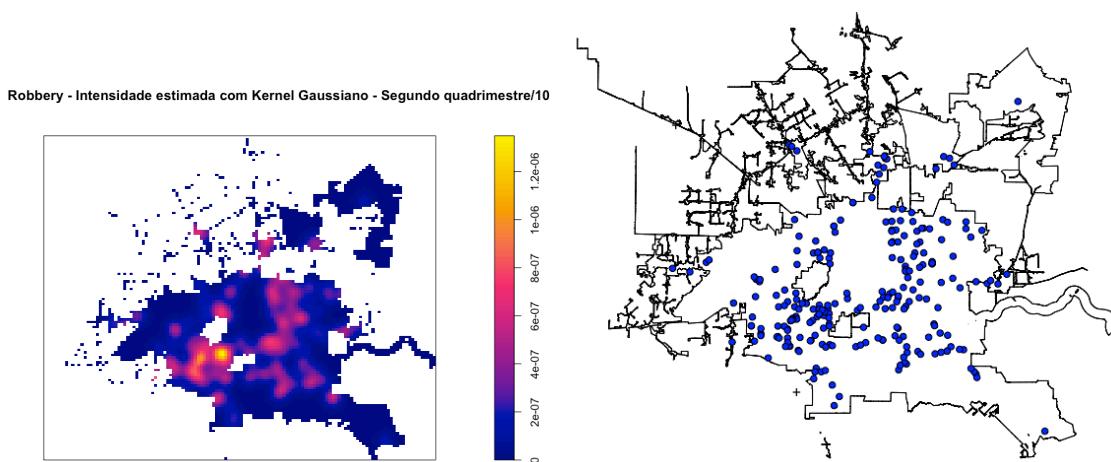
RAPE – Primeiro e segundo quadrimestre de 2010



ROBBERY – Primeiro e segundo quadrimestre de 2010

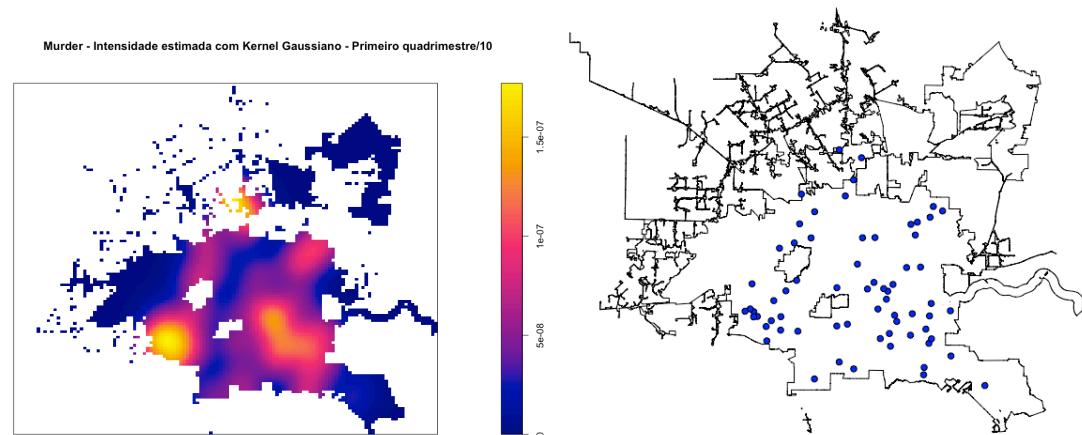


Robbery - Ocorrências no segundo quadrimestre/2010

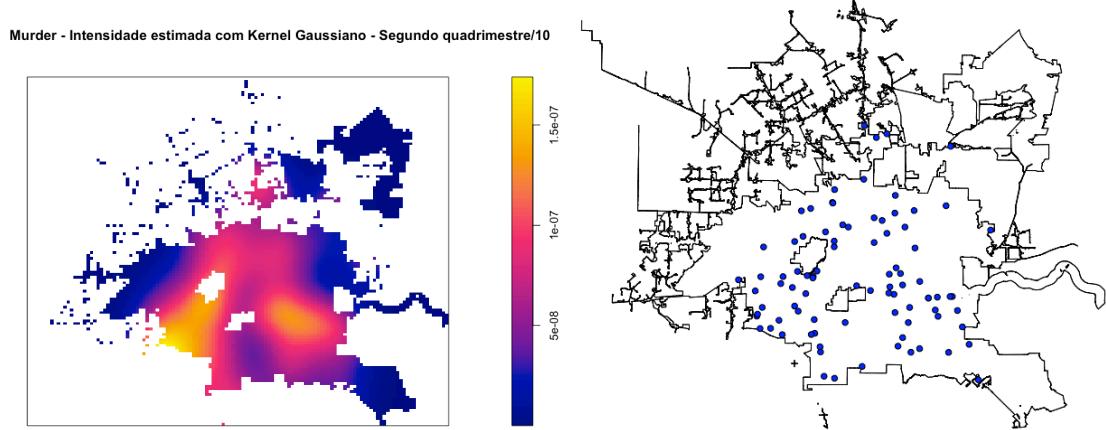


MURDER – Primeiro e segundo quadrimestre de 2010

Murder - Ocorrências no primeiro quadrimestre/2010



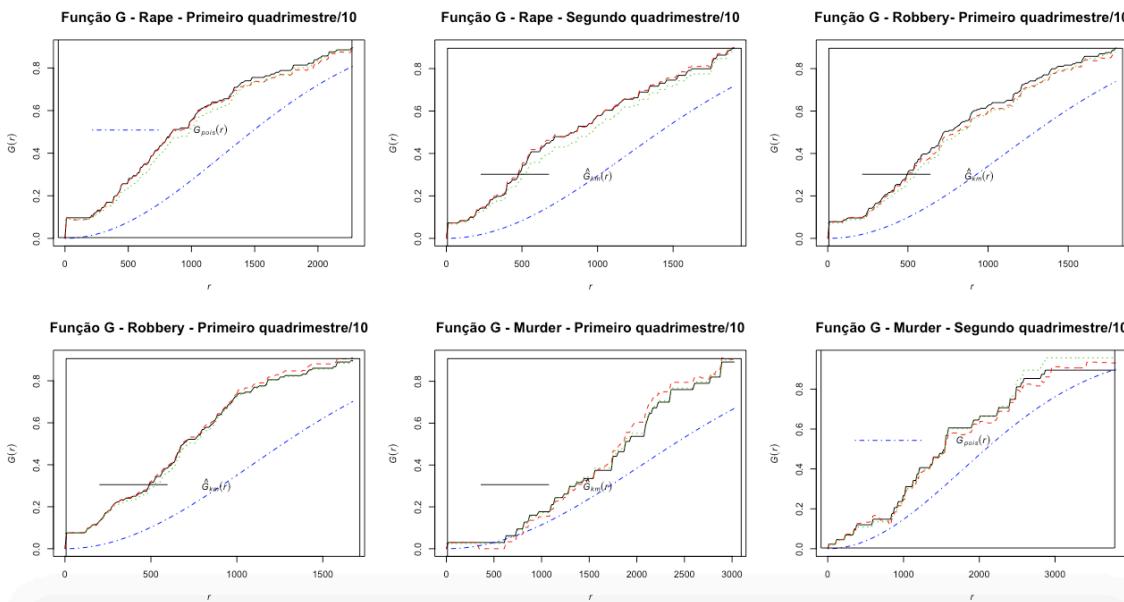
Murder - Ocorrências no segundo quadrimestre/2010



Por meio da análise e comparação entre os mapas, é possível observar os tons mais amarelados justamente onde estão concentrados os pontos azuis.

ESTIMAÇÃO DE EFEITOS DE SEGUNDA ORDER

Vamos analisar agora os efeitos de segunda ordem, onde foi aplicada a função G e logo após a geração dos gráficos a seguir que nos ajudarão a identificar uma autocorrelação espacial comparando com a curva que considera um completa aleatoriedade espacial.



É possível observar que há forte indicação de padrão de agrupamento dos pontos, levando em consideração as curvas calculadas da função G estarem localizadas acima das linhas aleatórias geradas pela poisson.

TESTE FORMAL CLARK-EVANS

Para verificação dessas hipóteses, optou-se ainda em realizar o teste formal de Clark-Evans, considerando com hipótese alternativa o agrupamento. Serão apresentados a seguir o p-valor calculado nesses testes, se menores de 0.05 se confirma a rejeição da hipótese H₀, ou seja, se aceita a hipótese H₁ de que há agrupamento.

```
Clark-Evans test  
No edge correction  
Z-test  
  
data: Houstonppp_rape1  
R = 0.81559, p-value = 1.432e-06  
alternative hypothesis: clustered (R < 1)  
  
Clark-Evans test  
No edge correction  
Z-test  
  
data: Houstonppp_rape2  
R = 0.8208, p-value = 9.555e-07  
alternative hypothesis: clustered (R < 1)  
  
Clark-Evans test  
No edge correction  
Z-test  
  
data: Houstonppp_robbery1  
R = 0.82452, p-value = 1.779e-07  
alternative hypothesis: clustered (R < 1)  
  
Clark-Evans test  
No edge correction  
Z-test  
  
data: Houstonppp_robbery2  
R = 0.73752, p-value = 4.717e-15  
alternative hypothesis: clustered (R < 1)  
  
Clark-Evans test  
No edge correction  
Z-test  
  
data: Houstonppp_murder1  
R = 0.85226, p-value = 0.01035  
alternative hypothesis: clustered (R < 1)  
  
Clark-Evans test  
No edge correction  
Z-test  
  
data: Houstonppp_murder2  
R = 0.91565, p-value = 0.06505  
alternative hypothesis: clustered (R < 1)
```

CONCLUSÃO

Com o estudo realizado, foi possível concluir de que os dados de crimes em Houston, na janela sugerida, na grande maioria não estão aleatoriamente distribuídos. Apenas o cenário de *murder* do segundo quadrimestre apresentou p-valor acima de 0.05 o que evidencia um padrão de completa aleatoriedade espacial.

Optou-se em não calcular o efeito de segunda ordem utilizando envelopes devido a exigência computacional que esses testes demandam, e devido a prévia confirmação do padrão de agrupamento verificado nos testes anteriores.